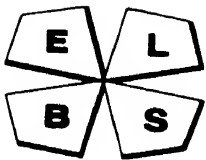


Muir's Textbook of Pathology

Eleventh Edition

Edited by J. R. Anderson

B.Sc., M.D., F.R.C.P.(Glas.), B.E., F.R.C.P.(Lond.), F.R.C.Path., F.R.S.(Edin.)
Professor of Pathology, University of Glasgow



**The English Language Book Society
and
Edward Arnold (Publishers) Ltd**

Preface



It has long been appreciated that the undergraduate course in medicine cannot do more than lay the basis for further training, and that the newly-qualified doctor must undergo a further period of general training, followed by the appropriate specialist or vocational training, before commencing independent practice. There are, however, wide differences of opinion on what the undergraduate student should be expected to know, and this is nowhere better illustrated than by the variations in the time allocated to pathology in the curricula of medical schools. Clearly the authors of textbooks of pathology are faced with a problem, both in the selection of topics to be included and in the depth of treatment of each topic. The purpose of this book is to provide a text suitable for both undergraduate students and for doctors training or practising in various specialties, including junior trainees in pathology. It contains more than will be assimilated by most undergraduate students during their formal course in pathology, and students using it should receive guidance from their teachers on which topics deserve their closest attention. I hope it will continue to be of use throughout the clinically-oriented part of the curriculum and during subsequent training.

A brief introductory chapter defines pathology, explains its central position in medical education, and describes its importance in patient care and in the advancement of medical knowledge. The remainder of the book is divided into two main sections 'general' and 'systematic' pathology.

The twelve chapters on 'general' pathology describe pathological processes of fundamental importance—mechanisms and effects of cell injury, the inflammatory response to injury, healing and repair, the physiology of the immune response and its beneficial and harmful consequences, infection and host-parasite relationships, local and general disturbances of blood flow, and the causation, types and behaviour of tumours. It is in some of these basic processes that advance has been most rapid, and indeed this has necessitated production of this edition

earlier than had been intended. The accounts of the cellular basis of immune responses and of the causation of tumours have been very largely rewritten and most of the other chapters have been changed considerably. These basic processes are applicable to a wide range of species and although they are, where possible, illustrated by human material, much of the text is based on the results of experimental work. The size of the general chapters has been determined not just by the clinical importance of the physiological and pathological processes they describe, but also by the amount of firm information that can usefully be imparted to students and trainee doctors. For example, the mediators of the inflammatory reaction and the basis of neoplasia are both subjects of considerable practical importance, but knowledge on them is still very limited and larger accounts would, I believe, be more likely to confuse than help the student. By contrast, the immunity system and its abnormalities are still relatively unimportant as the basis of primary illness in man, and yet they merit detailed consideration because they continue to be the subject of rapid scientific advance. In spite of our efforts at brevity, it has not been possible to avoid some increase in the length of the general section, which I hope will be of use to medical students and to biological scientists in general.

The 'systematic' section consists of fourteen chapters, each devoted to the more important diseases of man which affect a particular organ or system—the heart, lungs, blood and haemopoietic tissues, alimentary system, etc. Emphasis has been placed on the aetiology of those diseases, and their structural changes and effects on function, together with brief clinico-pathological correlations. Every effort has been made to update these chapters. A section has been added on dental and related oral pathology, while the disorders of the male and female reproductive systems have been rewritten as separate chapters with a brief account of sexually-transmitted diseases placed appropriately between them. Multi-authorship has been of particular value in the systematic chap-

Contents

<i>Chapter</i>	<i>Revised by</i>	
1 Introduction	<i>J. R. Anderson</i>	1
 General Pathology		
2 Cell Damage	<i>R. B. Goudie</i>	7
Section on the Effects of Ionising Radiation	<i>J. Stewart Orr</i>	
3 Inflammation	<i>J. R. Anderson</i>	43
4 Healing (Repair) and Hypertrophy	<i>Mary E. Catto</i>	77
5 Immunophysiology: The Immune Response	<i>R. B. Goudie and J. R. Anderson</i>	102
6 Immunopathology	<i>J. R. Anderson</i>	141
7 Host-Parasite Relationships	<i>J. R. Anderson</i>	174
8 Types of Infection	<i>J. R. Anderson</i>	193
9 Disturbances of Blood Flow and Body Fluids	<i>J. R. Anderson</i>	226
10 Miscellaneous Tissue Degenerations and Deposits	<i>J. R. Anderson</i>	269
11 Tumours I. General Features, Causation and Host Reactions	<i>A. J. Cochran</i>	290
12 Tumours II. Epithelial Varieties and Modes of Spread	<i>Bernard Lennox</i>	322
13 Tumours III. Other Varieties	<i>Bernard Lennox</i>	340
 Systematic Pathology		
14 Blood Vessels and Lymphatics	<i>J. R. Anderson</i>	360
15 The Heart	<i>J. R. Anderson</i>	396

16	Respiratory System	<i>Donald Heath and J. M. Kay</i>	432
17	The Blood and Bone Marrow	<i>John Dagg and F. D. Lee</i>	504
18	Lympho-reticular Tissues	<i>F. D. Lee and J. R. Anderson</i>	560
19	Alimentary Tract	<i>F. D. Lee</i>	587
	Section on Oral Pathology	<i>D. G. MacDonald</i>	
20	Liver, Biliary Tract and Exocrine Pancreas	<i>R. N. M. MacSween</i>	661
21	The Nervous System	<i>J. Hume Adams</i>	724
	Section on the Eye	<i>W. R. Lee</i>	
22	Urinary System	<i>J. R. Anderson and I. A. R. More</i>	805
23	Locomotor System	<i>Mary E. Catto</i>	874
24	Female Reproductive Tract	<i>H. Fox</i>	942
	Section on the Breast	<i>A. T. Sandison</i>	
25	Male Reproductive System	<i>Bernard Lennox</i>	987
26	The Endocrine System	<i>A. T. Sandison and J. R. Anderson</i>	1005
	Section on the Thyroid	<i>R. B. Goudie</i>	
	Section on the Pancreas	<i>R. N. M. MacSween</i>	
27	The Skin	<i>A. McQueen</i>	1049
	Index		1086

Introduction

What is pathology?

Pathology is the study of disease by scientific methods. Disease may, in turn, be defined as an abnormal variation in the structure or function of any part of the body. There must be an explanation of such variations from the normal in other words, diseases have causes, and pathology includes not only observation of the structural and functional changes throughout the course of a disease, but also elucidation of the factors which cause it. It is only by establishing the cause (*aetiology*) of a disease that logical methods can be devised for its prevention or cure. **Pathology may thus be described as the scientific study of the causes and effects of disease.**

Methods used in pathology

These include (*a*) **histology and cytology**, in which the structural changes in diseased tissues are examined by naked-eye inspection, or by light and electron microscopy of tissue sections or smears; (*b*) **biochemistry**, in which the metabolic disturbances of disease are investigated by assay of various normal and abnormal compounds in the blood, urine, etc.; (*c*) **microbiology**, in which body fluids, mucosal surfaces, excised tissues, etc., are examined by microscopical, cultural and serological techniques to detect and identify the micro-organisms responsible for many diseases.

These methods may be applied to the study of individuals suffering from a disease, and to animals in which a model of the disease occurs naturally or has been induced experimentally. The development of special techniques to investigate some types of disease has led to further specialisation in pathology. For example, the diagnosis of disorders of the blood involves various quantitative tests on, and morphologi-

cal examination of, the cells of the blood and haemopoietic tissue, assay of the factors involved in clotting, investigation of the metabolism of iron, vitamin B₁₂, etc., the detection of abnormal antibodies to cells of the blood and blood group serology. The many techniques involved have required the establishment of **haematology** laboratories: application of techniques to determine chromosome anomalies has led to the establishment of **cytogenetics** laboratories, and microbiology has divided into **bacteriology** and **virology**. Finally, **immunology**, a subject of enormous interest in biology and of increasing clinical significance, now requires special laboratory facilities. It will be apparent that pathology covers a wide spectrum of techniques, both in the diagnosis of patients and in research into the causes of various diseases. The relative importance of the several branches of pathology varies for different types of disease. In some instances, for example in diabetes mellitus, biochemical investigations provide the best means of diagnosis and are of the greatest value in the control of therapy. By contrast, recognition of the nature of many diseases, for example tumours, and so the choice of the most appropriate therapy, depend very largely on examination of the gross and microscopic features. For most diseases, diagnosis is based on a combination of pathological investigations. To give an example, biochemical tests may indicate that a patient is suffering from impairment of renal function, but the nature of the renal disease responsible for this commonly requires removal of a piece of renal tissue for histological examination (*renal biopsy*). Another example is provided by the condition of anaemia, which may have many causes. The changes in the cells of the blood and the bone marrow may suggest deficiency of a factor essential for erythropoiesis, and biochemical and physiological tests are then indicated to con-

2 Introduction

firm the deficiency, e.g. of vitamin B₁₂ or folic acid. Alternatively, anaemia may result from blood loss and this may be due to a structural lesion of the gastro-intestinal tract or of the endometrium, diagnosis of which may require histological examination.

The hospital pathologist is becoming much more clinically orientated. He must co-operate closely with clinicians, not only in diagnosis, but also by applying his skills to assessment of the effects of treatment, e.g. by examination of multiple biopsies of cancers and other lesions, removed serially during the course of treatment. He must also monitor patients for unwanted effects of treatment, e.g. the harmful effects of some drugs on the cells of the liver, kidney or haemopoietic tissue.

Finally, it is important to emphasise the continuing value of the clinical necropsy. In the past, when diagnostic procedures were relatively limited and primitive, a high proportion of diagnoses were made in the post-mortem room. In many cases, the more sophisticated diagnostic procedures now available have not diminished the value of necropsy, even in hospitals providing a very high standard of patient care (Cameron, 1978). The important role of post-mortem examination in elucidating the natural history of disease processes is well illustrated by the extensive studies of Willis (1973) on the spread of tumours within the body. This role of the necropsy is still important, for it is revealing the changes in the patterns of many diseases, and also new and unwanted effects, resulting from use of the ever-increasing number and variety of powerful drugs and therapeutic procedures available to the clinician.

Why learn pathology?

Most medical students are not going to become pathologists. It is nevertheless essential that the medical school curriculum should include a course of pathology which provides a clear account of the causes, where these are known, and of the pathological changes, of the more important diseases. Most disease processes bring about structural changes and these usually provide a logical explanation for the symptoms and signs and commonly also for the biochemical changes. A basic knowledge of the pathological processes of disease thus aids the

doctor in the correct interpretation of the clinical features of the patient's illness. This applies not only to the clinical diagnostician but also to the surgeon who must recognise the nature of the structural changes exposed at operation and act accordingly, and to the radiologist who must be familiar with the structural changes of diseases in order to interpret the shadows they cast on an x-ray film. To the research worker, histopathology and electron microscopy are superb techniques, both can be adapted to enzymic and other chemical investigations (histochemistry), including immunohistological techniques which make use of the exquisite specificity of antigen-antibody reactions to detect tissue and cell constituents and abnormal substances (see Fig. 22.20, p. 820 and Fig. 26.1, p. 1005).

Accordingly, pathology is of central importance to the medical student, regardless of the branch of medicine he intends to pursue.

How to learn pathology

Pathology is no exception to the general rule that learning is dependent mainly on the student's own effort. Most medical schools provide lectures and or small-group tutorials, demonstrations and practical classes in pathology, but self-education by reading, preferably supplemented by audio-visual aids, is essential. The student should also take full advantage of opportunities to compare the clinical features of patients' illnesses with the underlying pathology. Clinico-pathological conferences on selected cases, held for teaching purposes, are helpful but one of the best places to see pathology and to compare the clinical features of disease with the pathological changes is the post-mortem room. A well-conducted necropsy, presented jointly by a clinician who cared for the patient and the pathologist performing the necropsy, is still unsurpassed as a teaching method. Students should also gain experience by following the progress of the patients they examine, noting the results of laboratory investigations and where possible examining the lesions removed surgically or revealed at necropsy.

Pathology in the medical curriculum

There is a logical sequence in the pattern of teaching of most medical schools. After courses

in the basic sciences—chemistry, physics, biology often provided before starting at medical school, the student is introduced to normal human structure (anatomy and histology) and function (physiology and biochemistry), followed by courses in pathology (the causes, features and effects of diseases) and pharmacology, and finally concentrates on the clinical subjects, i.e. the diagnosis and treatment of patients. Classically, the subjects are dealt with on a broad front. For example, the courses in anatomy, etc. deal with the whole of the body. In many medical schools, this policy has been replaced by what is variously termed 'integrated', 'topic' or 'systems' teaching, in which each of the body's major systems (cardiovascular, alimentary, respiratory, etc.) is the subject of a teaching course provided by a multi-disciplinary team. Thus the course on, say, the alimentary system will include its anatomy, physiology, biochemistry, pathology, pharma-

cology and clinical aspects. Each method has its advantages, but it has become abundantly clear that the second method requires considerable organisation, and good co-operation between departments in the preparation and delivery of the course on each system. At present, there is a tendency to revert to the classical type of curriculum, or to compromise between the two.

One of the great advantages of a course in pathology, spanning the gap between the pre-clinical and clinical subjects, is that it provides the student, in the early part of his hospital experience, with a basic knowledge of the diseases he is likely to encounter most often in the wards and clinics. By contrast, the integrated course must either be brief and intensive or must extend over much of the curriculum, with the result that some systems come very late, leaving little time for their personal clinical study by the student.

Pathological processes

It was first pointed out by Virchow that all disturbances of function and structure in disease are due to cellular abnormalities and that the phenomena of a particular disease are brought about by a series of cellular changes. Pathological processes are of a dual nature, consisting firstly of the **changes of the injury** induced by the causal agent, and secondly of **reactive changes** which are often closely similar to physiological processes. If death is rapid, as for example in cyanide poisoning, there may be little or no structural changes of either type. Cyanide inhibits the cytochrome-oxidase systems of the cells and thus halts cellular respiration before histological changes can become prominent. Similarly, blockage of a coronary artery cuts off the blood supply to part of the myocardium and death may result immediately from cardiac arrest or ventricular fibrillation. When this happens, no structural changes are observed in the myocardium. If, however, the patient survives for some hours or more, the affected myocardium shows changes which occur subsequent to cell death and the lesion becomes readily visible both macroscopically (Fig. 15.9, p. 404) and microscopically (Fig. 2.5, p. 11).

Reactive changes may be exemplified by enlargement of the myocardium in the patient with high blood pressure (Fig. 4.31, p. 100). In this condition, there is an increase in the resistance to blood flow through the arterioles and consequently the normal rate of circulation can be maintained only by a rise in blood pressure. Reflex stimulation of the heart results in more forcible contractions of the left ventricle, and in accordance with the general principle that increased functional demand stimulates enlargement (**hypertrophy**) and/or proliferation (**hyperplasia**) of the cells concerned, the myocardial cells of the left ventricle increase in size. Although part of a disease state, the reactive hypertrophy of the myocardium in hypertension is closely similar to the physiological hypertrophy of the skeletal muscles in the trained athlete. To give another example, the invasion of the body by micro-organisms, in addition to causing injury, stimulates reactive changes in the lymphoid tissues, with the development of immunity. The distinction between the changes due to injury and those due to reaction are not usually so well defined as in the above examples. In many instances where cell injury

4 Introduction

persists without killing the cells, the cytological changes are complex and those due to injury often cannot be distinguished from those due to reaction. Some examples of the various types of cell injury and reaction are provided in Chapter 2.

In order to facilitate the understanding of pathological processes, it is helpful to group together those which have common causal factors and as a consequence exhibit similarities in their structural changes. For example, bacterial infections have certain features in common, and may with advantage be further sub-divided into acute and chronic infections. The features and behaviour of neoplasms (tumours) are sufficiently similar to classify most tumours into two categories, benign and malignant, and to provide a general account of each group. The changes resulting from a deficient blood supply are similar for all tissues. Accordingly, the next twelve chapters of this book are of a general nature and deal with the commoner pathological processes. The remaining chapters are systematic and go on to describe the special features of disease processes as they affect the various organs and systems.

The causes of disease

Causal factors in disease may be genetic or acquired. **Genetically-determined disease** is due to some abnormality of base sequence in the DNA of the fertilised ovum and the cells derived from it, or to reduplication, loss or misplacement of a whole or part of a chromosome. Such abnormalities are often inherited from one or both parents. **Acquired disease** is due to effects of some environmental factor, e.g. malnutrition or micro-organisms. Most diseases are acquired, but very often there is more than one causal factor and there may be many. Genetic variations may influence the susceptibility of an individual to environmental factors. Even in the case of infections, there is considerable individual variation in the severity of the disease. Of the many individuals who become infected with poliovirus, most develop immunity without becoming ill; some have a mild illness and a few become paralysed from involvement of the central nervous system (Fig. 21.44, p. 759). This illustrates the importance of host factors as well as causal agents. Spread of tuberculosis is favoured by poor personal

and domestic hygiene, by overcrowding, malnutrition and by various other diseases. Accordingly, disease results not only from exposure to the major causal agent but also from the existence of **predisposing or contributory factors**.

Congenital disease. Diseases may also be classified into those which develop during fetal life (congenital) and those which arise at any time thereafter during post-natal life. Genetically-determined diseases are commonly congenital, although some present many years after birth, a good example being adenomatosis (polyposis) coli, which is due to a dominant abnormal gene (see below) and consists of multiple tumours of the colonic mucosa, appearing in adolescence or adult life (Fig. 19.79, p. 682). Congenital diseases may also be acquired, an important example being provided by transmission of the virus of rubella (German measles) from mother to fetus during the first trimester of pregnancy. Depending on the stage of fetal development at which infection occurs, it may result in fetal death, or involvement of various tissues leading to mental deficiency, blindness, deafness or structural abnormalities of the heart. The mother may also transmit to the fetus various other infections, including syphilis and toxoplasmosis, with consequent congenital disease. Ingestion of various chemicals by the mother, as in the thalidomide disaster, may induce severe disorders of fetal development and growth. Another cause of acquired congenital disease is maternal fetal incompatibility. Fetal red cells exhibiting surface antigens inherited from the father may enter the maternal circulation and stimulate antibody production; the maternal antibody may pass through the placenta and react with the fetal red cells, causing a haemolytic anaemia.

Genetically-determined disease

As already mentioned, this results from abnormalities in the DNA which forms the genome. In some instances the abnormality consists of gain or loss of a whole chromosome or of part of a chromosome. Such gross abnormalities can now be detected by cell culture techniques. Most of them probably arise by non-disjunction of chromosomes in the meiosis which precedes germ-cell formation, and only a few appear to be compatible with life, e.g. an addi-

tional chromosome 21, which is the usual cause of Down's syndrome (mongolism).

A very large number of diseases result from the inheritance of an abnormal (mutant) gene, or combination of genes, from one or both parents. The development of abnormal genes (**mutation**) can be provoked by irradiation, mutagenic chemicals and probably by viruses, but in most instances the cause of mutations in man remains unknown. Examples of the many conditions resulting from an abnormal gene are colour blindness, albinism, haemophilia, sickle-cell anaemia, dystrophia myotomica and polyposis coli. The abnormal gene may be dominant, i.e. may induce an abnormality in spite of the presence of a normal corresponding gene from the other parent, or it may be recessive, i.e. causing disease only in the absence of a corresponding normal gene. The latter circumstance arises most usually in abnormalities of genes on the X chromosome, males being thus affected (Fig. 17.50, p. 585), or from the presence of two abnormal corresponding genes, one from each parent, the likelihood of which is enhanced by inbreeding.

In addition to those diseases due to mutations or recognisable chromosomal anomalies, there are many which show a **familial tendency**, but in which the mode of inheritance has not been elucidated. Examples include diabetes mellitus, chronic thyroiditis (see (6) below) and some of the commoner cancers, e.g. of the breast and of the bronchus. It is likely that both genetic and environmental factors are of causal importance in these conditions.

Acquired disease

The major causal factors may be classified as follows:

(1) **Deficiency diseases.** Inadequate diet still accounts for poor health in many parts of the world. It may take the form of deficiency either of major classes of food, usually high-grade protein, or of vitamins or elements essential for specific metabolic processes, e.g. iron for haemoglobin production. Often the deficiencies are multiple and complex. Disturbances of nutrition are by no means restricted to deficiencies, for in the more affluent countries obesity, due to overeating, has become increasingly common, with its attendant dangers of high blood pressure and heart disease.

(2) **Physical agents.** These include mechanical injury, heat, cold, electricity, irradiation and rapid changes in environmental pressure. In all instances, injury is caused by a high rate of transmission of particular forms of energy (kinetic, radiant, etc.) to or from the body. Important examples in this country are mechanical injury, particularly in road accidents, and burns. Exposure to ionising radiations cannot be regarded as entirely safe in any dosage. While radiation is used with benefit in various diagnostic and therapeutic procedures, any pollution of the environment with radioactive material is potentially harmful to those exposed to it and probably to subsequent generations.

(3) **Chemicals.** With the use of an ever increasing number of chemical agents as drugs, in industrial processes, and in the home, chemically-induced injury has become very common. The effects vary. At one extreme are those substances which have a general effect on cells, such as cyanide (see above) which causes death almost instantaneously, with little or no structural changes. Many other chemicals, such as strong acids and alkalis, cause local injury accompanied by an inflammatory reaction in the tissues exposed to them. A third large group of substances produces a more or less selective injury to a particular organ or cell type. Because of their important and complex metabolic activities, hepatocytes are injured by many chemical substances, including paracetamol and alcohol in high dosage. Many toxic chemicals or their metabolites are excreted by the kidneys, and because of their concentrating function the renal tubular epithelial cells are exposed to high levels of such substances. Accordingly toxic hepatic and renal tubular cell death are common. Fortunately both types of cell have a high regenerative capacity. Specific effects of chemicals are illustrated also by injury of neurones by overdosage of barbiturates and lung injury by paraquat (Fig. 16.37, p. 486).

(4) **Parasitic micro-organisms.** These include bacteria, protozoa, lower fungi and viruses. In spite of the advances in immunisation procedures and the extensive use now made of antibiotics, many important diseases still result from infection by micro-organisms, and the danger of widespread epidemics, e.g. of influenza and cholera, has been enhanced by air travel. The

disease-producing capacity of micro-organisms depends on their ability to invade and multiply within the host, and on the possibility of their transmission to other hosts. The features of the disease produced by infection depend on the specific properties of the causal organism. Bacteria bring about harmful effects mainly by the production of chemical compounds termed **toxins**, and the biological effects of these, together with the response of the host, determine the features of the disease. Viruses colonise host cells, and have a direct cytopathic effect: features of virus disease depend largely on which cells are colonised, the rate of viral replication, the nature of the cytopathic effect, and the response of the host. Of the protozoa, the malaria parasite is of enormous importance as a cause of chronic ill health in whole populations.

(5) **Metazoan parasites** are also an important cause of disease in many parts of the world. Hookworm infestation of the intestine and schistosomiasis are causes of ill health prevalent in many tropical countries.

(6) **Immunological factors.** The development of immunity is essential for protection against microbes and parasites. Harmful effects, both local and more widespread, can, however, result from the reaction of antibodies and lymphocytes with parasites, microbes and their toxic products. Also, the immunity system does not distinguish between harmful and harmless foreign antigenic materials, and injury may result from immune reactions to either. Such **hypersensitivity reactions** are numerous and complex. Local examples include hay fever, asthma and some forms of dermatitis, while hypersensitivity to many foreign materials, including penicillin and other drugs, sometimes

causes fatal generalised reactions. Hypersensitivity reactions may also result from the development of **auto-immunity** in which antibodies and lymphocytes develop which react with and injure normal cells and tissues: examples include chronic thyroiditis, commonly progressing to myxoedema, and the excessive destruction of red cells in auto-immune haemolytic anaemia.

In another group of disorders, the immunity system is deficient, and the patient lacks defence against micro-organisms: this may result from abnormalities of fetal development, as an effect of various acquired diseases, or may be induced by immuno-suppressive therapy.

(7) **Psychogenic factors.** The mental stresses imposed by conditions of life, particularly in technologically advanced communities, are probably largely responsible for three important and overlapping groups of diseases. First, acquired mental diseases such as schizophrenia and depression, for which no specific structural or biochemical basis has yet been found. Second, diseases of addiction, particularly to alcohol, various drugs and tobacco, these result in their own complications, for example alcohol predisposes to liver damage (Fig. 20.27, p. 682) and causes various neurological and mental disturbances, while cigarette smoking is the major cause of lung cancer (Fig. 16.45, p. 498) and chronic bronchitis, and is concerned also in peptic ulceration and coronary artery disease. The third group of diseases is heterogeneous, and includes peptic ulcer (Fig. 19.28, p. 609), high blood pressure and coronary artery disease (Fig. 15.12, p. 406). In these three important conditions, anxiety, overwork and frustration appear to be causal factors, although their modes of action are obscure.

References

Cameron, H. M. (1978). The autopsy: its role in modern hospital practice. *Investigative Cell Pathology*, 1, 297-300.

Willis, R. A. (1973). *The spread of tumours in the human body*, 3rd edn., pp. 400. Butterworths, Sevenoaks and London.

Cell Damage

All metabolic activities of the body are carried out and regulated by the cells of the tissues, and since the time of Virchow cell injury has been recognised as the central problem in pathology. It is clearly important to know what factors cause cell damage and how these lead to the cellular disorders which result in the states we recognise as diseases. Our knowledge of this large and important subject has been slow to develop due to the extremely complex inter-relationship of biological activities within the cell. Recently, however, there have been rapid advances, partly due to greatly improved techniques of biochemical analysis, fractionation of subcellular organelles and microscopy, and partly to the use of homogeneous experimental systems such as cultures of clones of genetically identical cells (bacterial or mammalian). Equally important has been the strategic use of cellular disorders in which the initial damage affects only one molecular constituent of the cell, thereby providing information about the normal and abnormal function of that constituent and its relationship to other activities of the cell. For this reason the most instructive forms of cellular damage are those due to abnormality of a single gene or to the effects of a selective poison.

Single gene defects. In at least one disease, sickle-cell anaemia, we probably know the entire sequence of events leading to cellular destruction. The sickle-cell abnormality is an inherited defect characterised clinically by rapid destruction of red blood cells. Apparently an error has occurred in copying one base in the sequence of 146 base triplets in the DNA constituting the gene for the beta polypeptide chain of haemoglobin. This error, transcribed through messenger RNA, results in the insertion of the amino acid valine instead of glutamic acid in position 6 from the N terminal end of the beta polypeptide chain and the shape of that end of the

chain is altered. This abnormality does not matter when haemoglobin is oxygenated, but as the haemoglobin molecule gives up oxygen it expands and the abnormal parts of the two beta chains come to project from the surface of the molecule and unite with alpha chains of adjacent molecules. Masses of long helical fibres of polymerised deoxygenated haemoglobin form and these impart to the red cells abnormal rigidity and a characteristic sickle shape which make them unduly prone to mechanical injury and subsequent phagocytosis within the spleen. It should be noted that, compared with most cells, red cells have a very simple structure and are easily obtained for study. Also, haemoglobin is one of the few proteins whose molecular configuration and amino acid sequences are known in detail, so that the substitution of a single amino acid is detectable and indicates, in turn, a single error in the DNA base sequence.

Many other defects of single genes result in clearly defined primary lesions. There may be absence of a particular enzyme with predictable effects such as accumulation of substrate and deficiency of the product of the missing enzyme, but these effects often lead to secondary, more complex abnormalities which may result also from various other primary defects. Genetic disorders are further considered on p. 14.

Experimental poisoning. Many genetic defects are incompatible with cell survival and so cannot readily be investigated. Accordingly, toxic chemicals have been widely used to investigate more severe forms of cell injury, for example the impairment of oxidative phosphorylation by fluoroacetate or cyanide (p. 9). Much information has been obtained by use of drugs known to have a specific effect on particular cellular functions: examples include actinomycin D which inhibits transcription of

DNA to mRNA, and colchicine which interferes with microtubular function. Various bacterial and other biological toxins have specific effects, for example cholera toxin disturbs the sodium pump and α -bungarotoxin from snake venom blocks acetylcholine receptors. It is possible, however, that such substances have additional effects on other cellular mechanisms. The action of some other poisons is indirect and less specific. Thus the classical experimental poison carbon tetrachloride is toxic to liver cells because it is metabolised by the microsomal enzyme P450 to produce free $\text{CCl}_3\cdot$ and CCl_2 radicals which lead to peroxidation of mRNA and of unsaturated fatty acids in cell membranes, and also to secondary disturbances of protein, fat and carbohydrate metabolism: electron microscopy shows damage to rough endoplasmic reticulum and later to other cellular organelles. Several other poisons cause similar effects on liver cells and, as in genetic abnormalities, it is evident that various different injuries cause trains of common secondary events, some of which lead to cell death.

In the following account only a few examples of cellular damage have been selected. The topic arises frequently in later chapters and our superficial treatment of this important subject is merely a reflection of our present basic ignorance.

It is convenient to consider the effects of cellular injury under two main headings: (1) **cell death** or *necrosis*, in which irreversible changes take place in the cell so that no further integrated function such as respiration or maintenance of selective membrane permeability is possible; (2) **lesser forms of damage** (sometimes described as degenerations) in which functions important for the economy of the cell or body are diminished or lost but in which integrated vital functions such as respiration and selective membrane permeability remain possible. Many lesser forms of cellular damage are reversible when the cause is withdrawn, for example the injury to neurones by therapeutic doses of anaesthetic drugs. Others, not resulting in cell death, are irreversible, e.g. radiation damage to chromosomes resulting in non-lethal genetic mutation.

Necrosis

Necrosis means the death of cells or groups of cells. It may occur suddenly, for example when cells are exposed to heat or toxic chemicals, or may be preceded by gradual and potentially reversible damage in which case the term *necrobiosis* is occasionally used.

Causes of necrosis

(a) **Marked impairment of blood supply**, usually due to obstruction of an end-artery (that is, one without adequate collaterals) is a common and important cause of necrosis, the necrotic area being known as an *infarct* (p. 246). Different cells can withstand the anoxia which results from ischaemia (impaired blood flow) for different periods, nerve cells, for example, die after only a few minutes, while fibrocytes survive much longer periods of anoxia.

(b) **Toxins**. Certain bacteria, plants, and animals such as snakes and scorpions, produce toxic organic compounds which even in very

small quantities can cause cell damage amounting to necrosis. Some toxins have identifiable enzyme activity; for example, the causal organism of gas gangrene, *Clostridium welchii*, forms a lecithinase which digests the lipoprotein of cell membranes. Diphtheria toxin appears to inhibit cellular protein synthesis by indirect interference with the transfer of aminoacyl-tRNA to ribosomes. Certain bacterial toxins, including those mentioned above, exert their effects not only locally, but are distributed via the bloodstream and other routes and so injure the cells of organs remote from the infection. The necrosis accompanying bacterial infection may be partly due to interference with the circulation brought about by toxic injury to the vascular endothelium with inflammation and sometimes thrombosis.

(c) **Immunological injury**. As will be described in Chapter 6, cell injury results in various ways from immune reactions. This is a feature of many infections, including tuberculosis in

which tuberculoprotein, a nontoxic product of the tubercle bacillus, evokes an immune reaction which, though protective in function, paradoxically leads also to necrosis of cells in the neighbourhood of the organism.

(d) **Infection of cells.** In certain infections, notably by viruses, the infecting agent proliferates within cells. Many viruses kill infected cells in tissue culture (cytopathic effect) and this is probably the cause of necrosis *in vivo* of the anterior horn cells of the spinal cord in poliomyelitis.

(e) **Chemical poisons.** Many chemicals in high concentration cause necrosis by non-selective denaturation of the cellular proteins (e.g. strong acids, strong alkalis, carboic acid, mercuric chloride). Cyanide and fluoroacetate are much more selective poisons and in low concentrations quickly cause cell death by interfering with oxidative production of energy from glucose, fatty acids and amino acids. As shown in Fig. 2.1 cyanide inhibits the enzyme cytochrome oxidase, thereby preventing the use

of oxygen, while fluoroacetate forms a powerful competitive inhibitor of the enzyme aconitase which normally converts citrate to isocitrate in the Krebs citric acid cycle. Necrosis of liver or other specialised cells results from the effects of poisoning, but in many instances the mode of interaction between poison and cell is obscure.

(f) **Physical agents.** Cells are very sensitive to heat and, depending on the type of cell, they die after variable periods of exposure to a temperature of 45°C. Cold is much less injurious and, provided certain precautions are taken, cell suspensions and even small animals can be frozen without being killed. Necrosis after frostbite is due to damage to capillaries, resulting in thrombosis which may even extend to the arteries. Radiation damage, also a cause of necrosis, is considered on p. 32. Mechanical trauma such as crushing may cause direct disruption of cells. Certain disorders of the nervous system are sometimes accompanied by necrotic lesions in the limbs; these 'trophic' lesions were previously attributed to an ill-defined effect of denervation on tissue nutrition but are now thought to result from mechanical trauma which occurs unnoticed because of sensory loss.

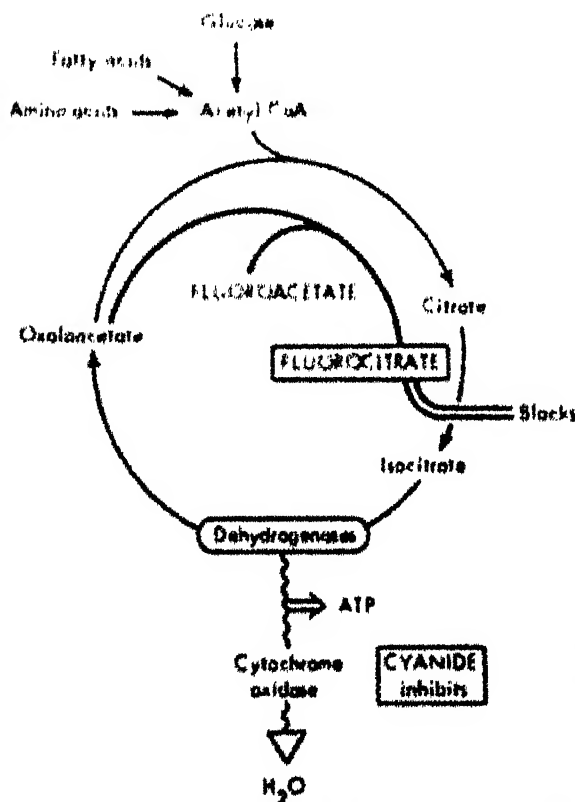


Fig. 2.1 The effects of fluoroacetate and of cyanide on cellular metabolism. Note that fluoroacetate is converted to fluorocitrate which inhibits conversion of citrate to isocitrate by aconitase.

The recognition of necrosis.

As a rule it is not possible to determine exactly when a particular cell becomes necrotic, i.e. when the disintegration of its vital functions has reached an irreversible stage. Many of the changes by which necrosis is recognised occur *after* cell death and are due to the secondary release of lytic enzymes normally sequestered within the cell, e.g. in the lysosomes; this process of *autolysis* is described below.

Necrosis of cell suspensions in tissue culture can be studied conveniently by observing changes of permeability of cell membranes to dyes such as neutral red or trypan blue. These dyes are normally excluded from the nucleus but when cells die, the nuclei become stained due to increased permeability of the membranes of the cell (Fig. 2.2). Alternatively, membranous components of the living cells may be labelled with radioisotopes such as ⁵¹Cr or ³²P; subsequent severe injury to the cell, probably lethal, is recognised by release of the radio-active label from the cells into the culture medium.

In organised tissues such as liver or kidney,

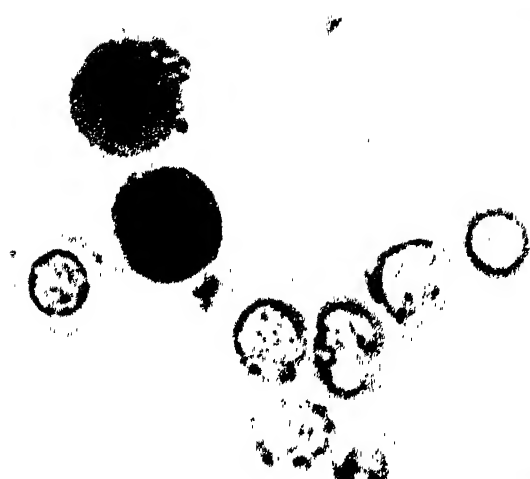


Fig. 2.2 A suspension of lymphocytes treated with cytotoxic iso-antibody and complement. Some of the cells have been killed, and have become stained by trypan blue dye present in the suspending fluid; other cells have survived and are unstained. $\times 1250$.

necrosis is usually recognised by secondary changes seen on histological examination. In preparations stained with haematoxylin and eosin, the nuclei may gradually lose their characteristic staining with haematoxylin so that the whole cell stains uniformly with eosin (Fig. 2.3), although the nuclear outline may persist; this change, the result of hydrolysis of chromatin within the cell after its death, is called **karyolysis**. Sometimes the chromatin of necrotic cells, especially those with already dense chromatin such as polymorphonuclear leucocytes, forms dense haematoxylinophilic masses (**pyknosis**) and these may break up (**karyorrhexis**) to form granules inside the nuclear membrane or throughout the cytoplasm (Fig. 2.4). In many necrotic lesions the outlines of swollen necrotic cells can be recognised but the cytoplasm is abnormally homogeneous or granular and frequently takes up more eosin than normal. In other tissues, e.g. the central nervous system, necrotic cells absorb water and then disintegrate, leaving no indication of the architecture of the original tissue; the lipids derived from myelin etc. persist in the debris of the necrotic tissue. The activities of certain enzymes, e.g. succinic acid dehydrogenase, diminish rapidly after cell death and appropriate tests provide useful indicators of recent tissue necrosis.

Electron microscopy of cells which have undergone necrosis shows severe disorganisa-



Fig. 2.3 Part of an infarct of kidney showing coagulative necrosis. A glomerulus and tubules are seen, but the nuclei have disappeared and the structural details are lost. $\times 172$.

tion of structure. Gaps are seen in the various membranes and abnormal polymorphic inclusions, presumably derived from membranes, lie in the ground substance. Fragmentation and vacuolation of endoplasmic reticulum and mitochondrial membranes precede the disap-

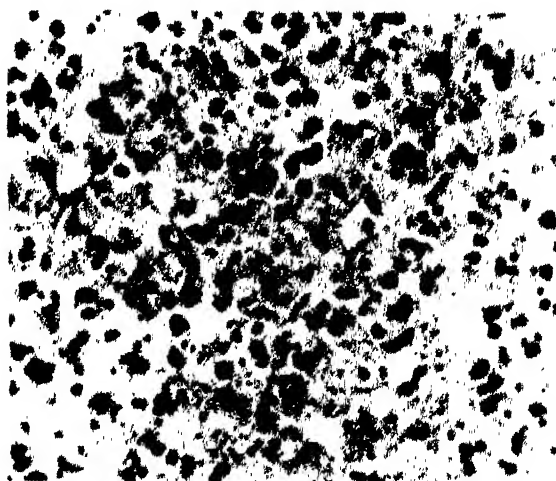


Fig. 2.4 Spreading necrosis with karyorrhexis in lymph node in typhoid fever. Note destruction of nuclei and numerous deeply-stained granules of chromatin. $\times 412$.



Fig. 2.6 Gangrene of toes.

action of phosphatases and proteolytic enzymes. The small molecules produced by hydrolysis of macromolecules lead to osmotic swelling of the necrotic cells and their organelles provided that the membranes are sufficiently intact.

It should be noted that when many polymorphonuclear leukocytes are present in necrotic tissue the enzymes from their abundant lysosomes may contribute to the hydrolysis of other cells. This is an important factor in the liquefaction of pus and in the softening seen in infected organs at necropsy.

If tissue is killed by heating, e.g. to 55°C, or by immersion in fixative such as formalin, the enzymes and other proteins are denatured and the histological features of necrosis attributable to autolysis do not develop. By contrast, if a piece of tissue is deprived of its blood supply by removal from the living body and kept at 37°C the development of autolysis can be observed, with marked osmotic swelling of membrane-bounded structures.

Two points of practical importance in the recognition of necrosis deserve emphasis. First, morphological signs of necrosis are not apparent until autolysis has developed in the necrotic tissue, and this takes 12–24 hours. Second, following death of the individual (somatic death), all cells of the body will in time die due to lack of blood supply and post-mortem autolysis will gradually take place. This is particularly marked in the parenchymal cells of the liver and kidney tubules and when seen at necropsy it may be mistaken for true necrosis, i.e. cell death occurring while the individual was still alive. This problem is of great importance in electron microscopy which shows fine structural evidence of necrosis and of post-mortem autolysis within a very short time.

Somatic death

Though not strictly related to cell necrosis, the interesting subject of somatic death (death of the individual) deserves some consideration.

Until recently, somatic death has been defined as complete and persistent cessation of respiration and circulation. For legal purposes the persistence of the state is arbitrarily taken as five or more minutes, by which time irreversible anoxic damage will have developed in the neurones of the vital centres. However, it is now possible to restore the circulatory and respiratory functions of heart and lungs in many cases of somatic death as defined above, and integrated function both of cells and of organs (excluding those of the central nervous system) can then continue for prolonged periods with the aid of special equipment. This fact is of great importance in obtaining organs for transplantation from cadaveric donors and a legal redefinition of somatic death in terms of extensive and irreversible brain damage is now necessary.

Effects of necrosis

By definition, necrotic cells are functionless. The effect of cell necrosis on the general well-being of the body accordingly depends on the functional importance of the tissue involved, the extent of the necrosis, the functional reserve of the tissue, and on the capacity of surviving cells to proliferate and replace those which have become necrotic. For example, splenectomy is compatible with good health in man (although it increases the risk of certain infections) and extensive splenic necrosis is apparently of little importance. By contrast, extensive necrosis of renal tubular epithelium results in the serious clinical condition of renal failure which is likely to be fatal unless the patient is kept alive (e.g. by haemodialysis) until there is regeneration of tubules by proliferation of surviving cells. Necrosis of a relatively small number of motor nerve cells may produce severe paralysis which persists because nerve cells cannot proliferate to replace those lost. Since myocardial cells have not only a contractile but also a conducting function, quite small necrotic lesions may result in striking alterations in the electrical activity of the heart.

The breakdown of necrotic cells results in

escape of their contents. Enzymes such as aminotransferases released into the plasma from necrotic liver or myocardial cells form the basis of clinical tests for necrosis in these tissues. It should be emphasised, however, that abnormal enzyme release occurs from cells with damage short of necrosis (e.g. in muscular dystrophy). In poisoning by alloxan, which kills the β cells of the pancreatic islets, discharge of stored insulin from the necrotic cells results in hypoglycaemia which may be fatal: those animals which survive develop diabetes from lack of insulin.

Reactions to necrosis

Neutrophil polymorphs frequently accumulate in small numbers around necrotic cells (Fig. 2.5). Occasionally infarcts and caseous lesions are invaded by large numbers of these cells and this leads to softening as already described. Such softening is a notable feature in a small proportion of myocardial infarcts (which usually show coagulative necrosis) and may lead to rupture of the heart; it is also common in tuberculosis of the lumbar vertebrae where the caseous material liquefies and tracks down beneath the psoas fascia to form a 'cold abscess' in the groin.

Individual cells killed by toxins rapidly undergo autolysis and are absorbed, especially when the circulation is maintained. They may be quickly replaced by proliferation of adjacent surviving cells. When a large mass of tissue undergoes necrosis, e.g. in an infarct, the necrotic material may be gradually replaced by ingrowth of capillaries and fibroblasts from the surrounding viable tissue so that a fibrous scar results. If this process is incomplete the necrotic mass becomes enclosed in a fibrous capsule, may persist for a long time, and may become calcified. Areas of necrotic softening in the brain are usually invaded by macrophages and eventually become cyst-like spaces containing clear liquid and surrounded by proliferated astroglia.

Old caseous lesions and necrotic fat have a marked affinity for calcium and frequently become heavily calcified.

Cell Damage Short of Necrosis

Many forms of injury may cause cellular abnormalities short of necrosis, which may seriously affect health. Such cellular abnormalities may be detected by impairment of a physiological activity such as conduction of a nerve impulse, by chemical or histochemical means (diminished or excessive enzyme activity or storage or depletion of a chemical substance), by structural abnormality revealed by microscopy of one kind or another, or by a combination of these methods. Some of these forms of cellular injury lend themselves to scientific study because they can be accurately, if arbitrarily, defined.

The following discussion deals with very heterogeneous topics. First we consider damage to the nucleus, membranes and organelles of the cell, then give examples of damage resulting in abnormal storage of metabolites. Next, the important problem of irradiation damage, both to single cells and cell populations, is discussed and finally, shrinkage of cells (**atrophy**) and alteration of cell structure to a form more resistant to injury (**metaplasia**).

Nuclear damage

The importance of nuclear damage depends on the fact that the cell nucleus contains the genetic information upon which all the vital activities of the cell ultimately depend. Indeed, as already explained, severe nuclear damage indicated by pyknosis and karyolysis are customarily taken as evidence of cell necrosis. It should be remembered, however, that red blood cells, although devoid of a nucleus, maintain selective membrane permeability, produce energy by anaerobic glycolysis and perform their vital specialised function of oxygen transport in the blood for over 100 days in man. Protozoa such as *Amoeba proteus* survive at least for several days following microsurgical removal of their nucleus: motility and phagocytic activity are arrested but these return, together with the ability to reproduce, when the nucleus from another amoeba is introduced. It is therefore clear that cells can survive despite total cessation of nuclear function; their metabolic versatility will, however, be greatly reduced and their ability to multiply lost.

Gene mutation

Perhaps the best understood form of nuclear damage is that due to irradiation or to mutagenic chemicals such as nitrogen mustards. These and other unidentified factors may result in errors in the sequence of purine and pyrimidine bases in DNA molecules. If the damage is sufficiently localised, e.g. affecting only one base, it is most unlikely to lead to an alteration in the nucleus demonstrable by available chemical or morphological techniques. Its presence may be inferred if there is a familial disease which can be shown to be due to genetic rather than purely environmental factors. Such diseases are essentially mediated by an abnormal gene resulting in incorporation of a 'wrong' amino acid at a functionally important part of a protein molecule, or by deletion of a gene with consequent absence of the protein. In most known examples, the affected proteins are enzymes, as in phenylketonuric oligophrenia which affects 1 in 10 000 of the population. It is due to deficiency in the liver cells of phenylalanine hydroxylase, which normally converts phenylalanine to tyrosine, and brain damage apparently results from raised blood and cerebrospinal fluid levels of phenylalanine and its metabolites. There are also many genetic abnormalities of haemoglobin (haemoglobinopathies p. 522), the sickle cell abnormality (p. 7) being a good example. Other carrier proteins, for example transferrin which binds iron, may also be defective. Genetic abnormalities of cell surface receptors, e.g. for low density lipoproteins (p. 30) have been described and there are probably errors in genes coding for the polypeptide chains of structural proteins such as collagen and also for proteins which regulate the activity of other genes. Many of the so-called inborn errors of metabolism are inherited as a recessive trait—i.e. are apparent only in the homozygous individual, although sometimes the symptomless heterozygote can be identified by demonstrating subnormal activity of the gene product. It should be noted that similar but not identical genetic diseases may result from mutations affecting different genes concerned in a particular function, for example the various factors involved in blood coagulation, or

by different mutations affecting a single gene and causing various degrees of functional impairment of the gene product. The influence of environmental factors in causing lesions may result in variations among individuals with an identical genetic abnormality, for example the severity of brain damage in phenylketonuric oligophrenia (see above) is influenced by the amount of phenylalanine in the diet.

Some genetic disorders are evident at birth (i.e. are **congenital**), particularly those resulting in abnormalities of physical development. Fig. 2.7 shows the masculinising effect of an excess of the androgenic steroid 17 α -hydroxyprogesterone on the development of the external genitalia of a female child. This is due to genetic deficiency of the adrenocortical enzyme which normally converts that substrate to a non-androgenic steroid in the biosynthesis of cortisol. Other genetic disorders become apparent when the demands of an independent existence after birth reveal a disorder of function such as renal failure, increased susceptibility to infection, or mental deficiency. Some genetic diseases are first recognised later in childhood or even in adult life because of the existence of metabolic pathways or functions peculiar to later periods of life or to the cumulative effects of factors which injure the genetically abnormal

cells. The number of genetic disorders of metabolism is increasing rapidly, although the biochemical basis of many has not yet been elucidated. In recent years it has become possible to diagnose certain genetic abnormalities by investigation of fetal cells obtained in aspirated amniotic fluid and this has sometimes provided the opportunity to terminate pregnancy before the fetus is viable.

In the above examples the mutation has occurred in a germ cell and is transmitted to the descendants of the affected individual, resulting in a familial pattern of disease. Mutation can also occur in cells other than germ cells—**somatic mutation**—but this is likely to be apparent only when the genetically altered cell proliferates abnormally to form a large family or clone of similar cells, e.g. when a tumour forms (pp. 319, 551).

An interesting abnormality encountered in many types of tumour is the synthesis of substances normally restricted to fetal life, e.g. α -fetoprotein (normally produced by fetal hepatocytes) by liver cell cancer in the adult. Some tumours produce compounds characteristic of other tissues, e.g. parathyroid hormone by cells of tumours arising in the bronchus. The mechanism of the underlying disorder of genetic control in such tumours is unknown.



Fig. 2.7 Developmental abnormality of external genitalia of newborn female child due to an inborn error of metabolism, androgen being synthesised in the adrenal cortex instead of cortisol. (Adrenogenital syndrome due to 21 hydroxylase deficiency.) (Professor M. A. Ferguson-Smith.)

Chromosomal abnormalities

Damage to the genetic apparatus more gross than that described above can sometimes be seen when the chromosomes of dividing cells are examined microscopically. An extra chromosome may be found, e.g. three instead of the normal pair of No. 21 chromosomes are commonly present in cases of Down's syndrome, (Fig. 2.8), but it is not understood how the chromosome abnormality leads to the physical and mental defects found in this condition. Total absence of one chromosome from all body cells is almost invariably incompatible with survival, with the notable exception of the Y sex chromosome, present in males (XY) but not in normal females (XX). Individuals with one X and no Y chromosome are phenotypically female but have a group of physical abnormalities including dwarfism and failure of ovarian development known as Turner's syndrome. Structural abnormality of chromosomes in the form of deleted portions



Fig. 2.8 Karyotype (trypsin - Leishman banded) of female infant with trisomic Down's syndrome. Three 21 chromosomes instead of the normal two. (Professor M. A. Ferguson-Smith.)

or added pieces derived from other chromosomes, or unusual shapes such as rings, are sometimes found and may be associated with characteristic clinical syndromes (Fig. 2.9). In a familial variety of Down's syndrome the two No. 21 chromosomes are normal but there is an abnormal No. 13 chromosome with an attached extra piece derived from a No. 21. This arises from a reciprocal exchange of fragments between two chromosomes during meiosis.

As would be expected, the gross chromosomal abnormalities described above, affecting all the cells of the body, lead to complex abnormalities since many genes must be involved. They arise during meiotic division of germ cells, the presence of an extra chromosome or absence of a chromosome being due to failure of separation of a pair of homologous chromosomes (**non-disjunction**); one of the resulting gametes will have an extra chromosome and

the other will be correspondingly defective. Structural chromosomal abnormality is due to chromosome breakage with re-arrangement of

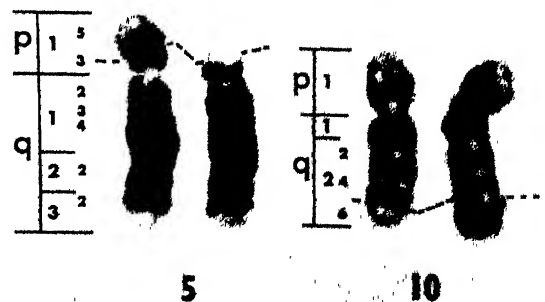


Fig. 2.9 Inherited structural chromosome abnormality in phenotypically normal individual. There is translocation of part of the short arm of a chromosome 5 to the long arm of chromosome 10 as shown by the banding pattern. (Professor M. A. Ferguson-Smith.)

fragments during repair. The reciprocal exchange of unequal fragments between non-homologous chromosomes (**translocation**) accounts for the occurrence of abnormally large or small chromosomes.

Abnormal chromosome numbers (**aneuploidy**) and structural chromosomal aberrations are invariably found in the cells of malignant tumours. The best known example of a consistent structural chromosomal aberration is the 'Philadelphia' chromosome, present in white and red blood cell precursors in the marrow in chronic myeloid leukaemia (a neoplastic proliferation of leukocytes). Radiation damage is known to cause chromosomal abnormalities in somatic cells and the frequency of chromosomal breakage has been used to assess exposure to radiation. Aneuploidy and structural abnormalities in individual chromosomes are also found in other circumstances, e.g. in thyroid epithelial cells which have proliferated following stimulation by pituitary thyrotrophic hormone, and in cells damaged by viruses.

Staining techniques which demonstrate structural transverse banding patterns are revealing less gross abnormalities and variations in chromosomes and are proving valuable in relating normal and abnormal genes to particular sites on individual chromosomes (**gene mapping**).

Nutritional nuclear damage

An interesting and clinically important form of nuclear damage is encountered in patients deficient in vitamin B₁₂ or folic acid. The nuclei are larger than normal but contain less than optimal amounts of DNA for cell division. The chromatin of the large nuclei is arranged in a fine threadlike fashion (Fig. 17.2, p. 505) compared with the condensed masses seen normally (Fig. 17.1, p. 505), and when mitosis occurs the chromosomes in affected individuals are longer and less tightly coiled than normal. These changes occur in many tissues but are best known in the precursors of red cells in the bone marrow which is said to exhibit **megaloblastic erythropoiesis**. In addition to nuclear enlargement in megaloblasts there is increased amount

of cytoplasm, cytoplasmic basophilia due to excessive RNA and premature haemoglobinisation as judged by the immature state of the nucleus.

The mechanism of these changes is incompletely understood. Folate plays an essential role in the synthesis of purine bases and thymine, and deficiency of these substances presumably impairs nucleic acid synthesis, especially DNA which, unlike RNA, contains thymine. This could explain the delay of DNA synthesis prior to cell division together with excessive cytoplasmic growth and premature haemoglobinisation. Vitamin B₁₂ is thought to influence nuclear structure by affecting folate metabolism (see Chanarin, 1979). Methyl folate is inactive in purine and thymine biosynthesis and one of the main functions of vitamin B₁₂ is the transfer of methyl groups from methyl folate for the synthesis of choline. Accordingly, when there is severe vitamin B₁₂ deficiency as in pernicious anaemia (p. 535), megaloblastic change occurs due to accumulation of methyl folate, and is reversed temporarily by the administration of folic acid, and permanently by life-long administration of vitamin B₁₂.

Toxic nuclear damage

Many drugs used in the treatment of cancer impair DNA replication either by combining directly with DNA (e.g. the alkylating agents nitrogen mustard and cyclophosphamide and certain antibiotics such as Actinomycin and Adriamycin) or by acting as analogues of normal metabolites and blocking the enzymes involved in nucleic acid synthesis. These include purine antagonists (mercaptopurine and its derivatives), pyrimidine antagonists (fluorouracil, cytarabine) and folic acid antagonists (methotrexate). The periwinkle alkaloid Vincristine damages the mitotic spindle. Amanitine (a toxic peptide from the mushroom *Amanita phalloides*) and D-galactosamine interfere with RNA synthesis in different ways and result respectively in fragmentation of nucleoli and in their disorganisation and replacement by small fibrillar bodies.

Damage to membranes and organelles

Electron microscopy reveals membranous structures which form the boundary wall around the cell and various compartments (organelles) within. The membranes are composed of lipoprotein or of phosphoglyceride bilayers containing enzymes, etc., and have the general properties of semipermeable membranes. The important definition of cellular compartments depends to a large extent on this property since soluble proteins of different types can thereby be sequestered within the cell; for example, hydrolases, potentially harmful to the cell, are confined within the lysosomes. As a result of the semipermeability of the membrane, the cell and its organelles tend to be subject to swelling and shrinkage depending on the relative osmotic pressures of the solutions in their various compartments and in the extracellular fluid. The membranes are not, however, inert but actively regulate the transport of crystalloids, including electrolytes, by enzymatic action which constantly modifies the chemical structure of the membrane and requires the provision of energy from ATP. Thus although K^+ and to a smaller extent Na^+ can passively diffuse through cell membranes, the intracellular concentration of K^+ is much higher, and of Na^+ lower, than that of the extracellular fluid; these differences are due to the outward 'pumping' of sodium by the cell membrane.

An additional important function of the membranes of the cell is that they form a cytoskeleton. This influences the shape of the cell and provides supporting structures for arrays of enzymes which form sequential functional units, such as those involved in the citric acid cycle and the flavoprotein and cytochrome systems of the cristae of the mitochondria.

Cell membranes

It has been shown by microsurgery that cells can survive incision of the surface membrane, and presumably self-sealing gaps develop in membranes when particulate material (e.g. nuclear fragments from normoblasts) is extruded from cells. However, most forms of reversible injury to surface membranes are not associated with demonstrable structural lesions. When antibody reacts with antigen associated with

the cell surface, complement may be activated (p.143) with formation of complexes which bind to and injure the cell membrane, producing blebs and holes in it (Fig. 2.10); cell death results from osmotic disturbances.



Fig. 2.10 Electron micrograph of part of the cell membrane of *Esch. coli* treated with antibody and complement, followed by treatment with trypsin. Activation of complement at the sites of antigen-antibody reaction has resulted in lesions—apparently holes—in the cell membrane, and these are accentuated by trypsin. $\times 140\,000$. (Dr. R. Dourmashkin.)

An indication of surface membrane dysfunction frequently encountered following anoxia and certain poisons is osmotic swelling of the cell with accumulation of water in the cytoplasm and resulting separation of organelles (Fig. 2.11 and 2.12). Such intracellular oedema, when reversible, may be the result of increased permeability of the surface membrane to sodium, or due to failure to remove sodium from the cell consequent upon diminished supply of ATP, or to poisoning of the enzymes involved in the sodium pump, e.g. by the drug ouabain. Increased cell volume leaves less room for extracellular fluid and this may impair the transport of metabolites between cells and the

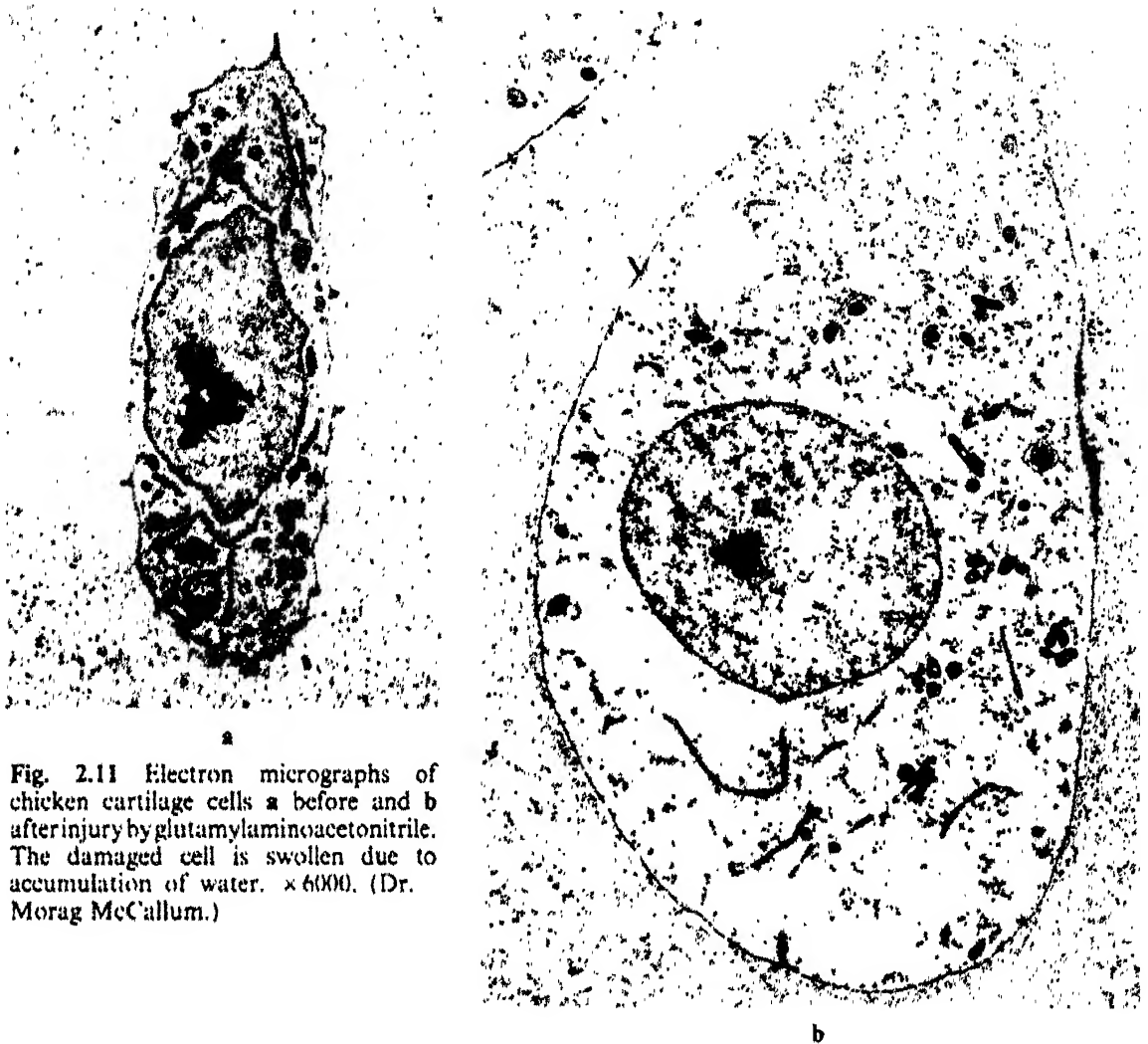


Fig. 2.11 Electron micrographs of chicken cartilage cells **a** before and **b** after injury by glutamylaminoacetonitrile. The damaged cell is swollen due to accumulation of water. $\times 6000$. (Dr. Morag McCallum.)

circulation, with further cell injury. When sodium is taken up by large numbers of cells, for example around extensive burns, there may be severe hyponatraemia (**sick cell syndrome**) which should not be treated by administering sodium.

The cell membrane controls the passage not only of water and inorganic ions, but also of organic molecules, into the cell cytoplasm. An inherited disorder characterised by failure of membrane transport of diamino acids, including lysine and ornithine, results in their presence in the urine because the renal tubular epithelium cannot re-absorb them: there is also a raised blood level of ammonia which is not converted to urea because ornithine does not enter hepatocytes in amounts adequate for the urea cycle.

The cell membrane also has an important

role in receiving information from the environment of the cell through specialised surface receptor molecules which can combine specifically with polypeptide hormones, antigens, etc. Diminished numbers of surface receptors, leading to altered cellular responses, may result from genetic abnormality (e.g. in hyperbeta-lipoproteinaemia, p. 30), blocking by poison (e.g. the acetylcholine receptor on skeletal muscle by curare) or from prolonged overstimulation (e.g. of the insulin receptor by chronic hyperinsulinaemia).

Damage to **desmosomes** (the adhesion points of cell membranes which bind epithelial cells together), apparently caused by antibody and complement acting on adjacent intercellular material, is seen in the skin disease pemphigus. The desmosomes of the stratified squamous epithelium of the skin and mucous membranes



Fig. 2.12 Ballooning of epithelial cells due to bacterial toxic injury in acute laryngitis. $\times 250$.

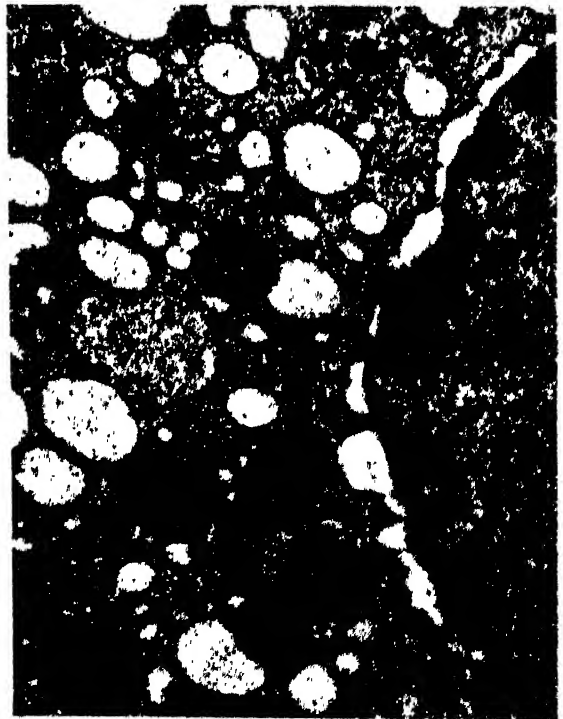
(the 'prickles' of the prickle cell layer) disappear and the resulting loss of cellular adhesion (acantholysis) is expressed in the formation of large intra-epidermal blisters (bullae) containing viable disaggregated prickle cells.

Endoplasmic reticulum

Loss of parallel arrays of endoplasmic reticulum and vacuolation due to accumulation of water within the membrane-lined spaces are frequently encountered as reversible lesions in anoxia and various poisonings, again presumably due to alterations in membrane permeability (Fig. 2.13). The swelling may be so severe in carbon tetrachloride poisoning as to give the cells a ballooned appearance. A marked increase of smooth endoplasmic reticulum bearing the mixed-function oxidases, e.g. cytochrome P450, is seen following the administration of certain drugs, notably phenobarbitone. This change is of special importance in



a



b

Fig. 2.13 Effect of carbon tetrachloride on the liver cells. **a** Part of normal mouse liver cell; note the regular parallel plates of granular endoplasmic reticulum and discrete clusters of ribosomes. **b** Part of a mouse liver cell 4 hours after oral administration of carbon tetrachloride. The plates of granular endoplasmic reticulum appear to have segregated into smaller oval vesicles from which many of the ribosomes have become detached and are dispersed singly in the cytoplasmic matrix. Mitochondria show no abnormality. $\times 23\ 000$.

the study of cellular damage since these enzymes are responsible for the initial stages in the metabolism of many drugs, some of which are detoxicated and rendered less effective while the oxidation products of others are more toxic than the parent substances.

Disaggregation of polyribosomes, presumably associated with decreased production of mRNA, has been noted in ischaemic cell damage and with certain poisons, some of which also cause loss of ribosomal particles from the rough endoplasmic reticulum. The resulting failure of protein synthesis has been corrected in experimental situations by the provision of a synthetic mRNA, but when the outlines of the ribosomal particles (as seen by electron microscopy) become indistinct, there is irreversible failure of protein synthesis.

Mitochondria

Diminished oxygen supply quickly interferes with the important mitochondrial function of oxidative phosphorylation—the production of high energy phosphate bonds in ATP by combination of oxygen with hydrogen through the flavoprotein-cytochrome enzyme systems. One

minute of ischaemia causes a ten-fold decrease in the ATP:ADP ratio. Mitochondrial function can be restored by a return of adequate oxygenation even after lethal changes have occurred elsewhere in the cell.

Anoxia and many poisons cause reversible osmotic swelling of mitochondria which gives the cell cytoplasm a swollen, cloudy and granular appearance in light microscopy (Fig. 2.14). This change, long known as 'cloudy swelling', is seen especially in metabolically active tissues such as liver, kidney and myocardium, and is exactly the same as that which develops within a few minutes of cessation of the circulation after death of the body or excision of the tissue. For this reason, pathological significance can be attached to its finding only if pieces of tissue small enough to be permeated rapidly are promptly placed in fixative. Isolated mitochondria can be made to swell and contract *in vitro* by adding calcium ions and ATP respectively to the medium in which the mitochondria are suspended.

Electron microscopy of injured mitochondria, in addition to showing in detail the site of swelling, reveals other abnormalities. A very early indication of anoxic damage is the dis-

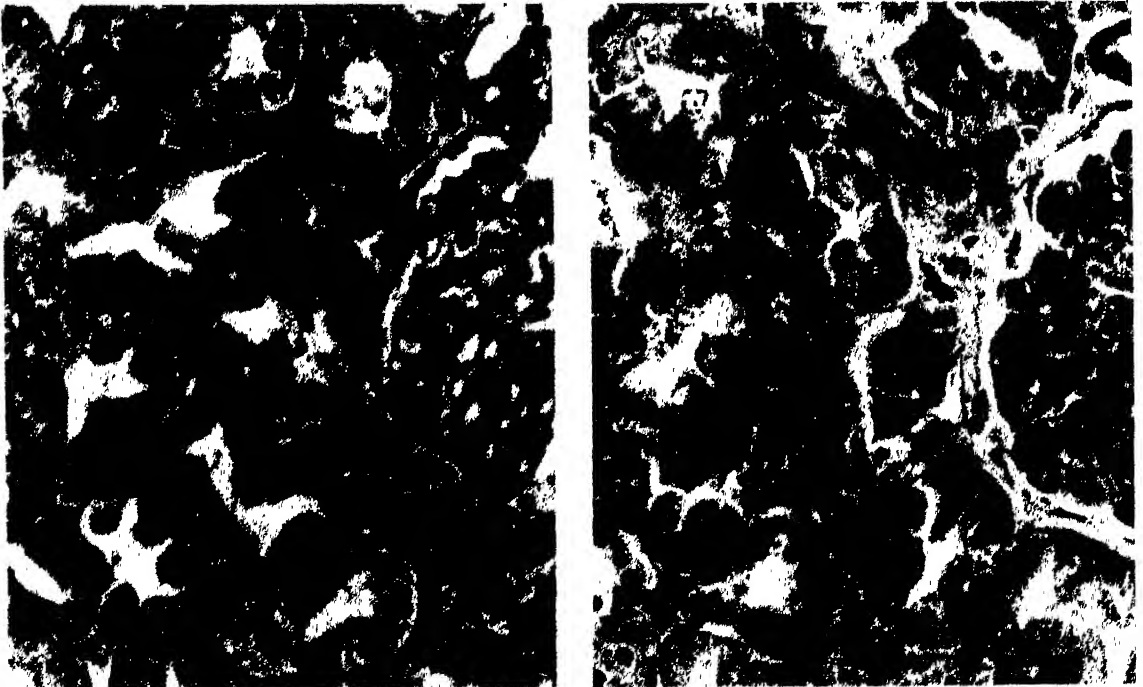


Fig. 2.14 *Left*, normal kidney. *Right*, cloudy swelling of renal tubular epithelium, showing cytoplasmic granularity. $\times 300$.

appearance of the dense granules occasionally seen in the matrix of normal mitochondria. These bodies are thought to represent lipid-bound calcium which accumulates under the influence of respiration and ATP. A later structural change, characteristic of mitochondrial damage following various types of cell injury, is development in the matrix of dense poorly defined aggregations of unknown composition. A second type of abnormal mitochondrial inclusion apparently composed of calcium (sometimes as hydroxyapatite) is found in very varied circumstances, e.g. in renal tubular epithelium when there is excessive calcium in the urine due to vitamin D poisoning or to hyperparathyroidism, and in myocardial mitochondria in

magnesium deficiency. Dense mitochondrial inclusions (Fig. 2.15) containing iron are found in the red cell precursors in sideroblastic anaemia in which there is impairment of iron utilisation (p. 540).

Lysosomes

These are cytoplasmic organelles limited by a single membrane. They contain various hydrolases which digest proteins, fats, carbohydrates, etc. Under the electron microscope they present very varied appearances and can be recognised with certainty only by histochemical demonstration of their acid hydrolase activity. The enzymes have a low pH optimum and become active when the membrane of the **primary lysosome** is altered, e.g. by fusion with a phagocytic vacuole to form a **secondary lysosome (phagosome, or phagolysosome)**.

Lysosomal hydrolases seriously impair the biochemical function and structure of subcellular particles *in vitro* and there has been much speculation on the importance of lysosomal damage in cell injury *in vivo* (see Dingle and Fell, 1973). Damage to the lysosomal membrane leading to release of lysosomal enzymes into the cytoplasm of living cells results in various degrees of cell damage up to necrosis. This happens in certain bacterial infections, e.g. by streptococci, and appears to result from the action of bacterial toxins. It is also encountered in hypervitaminosis A where it is attributed to the surfactant effect of the vitamin on the lysosomal membranes. Another example is the necrosis of macrophages which have ingested silica particles; some of the silica of the particles within phagolysosomes is converted to silicic acid and this forms hydrogen bonds with the phospholipids of the lysosomal membrane which then ruptures and releases the enzymes into the cytoplasm. Phagocytosis of monosodium urate crystals by neutrophil polymorphs in patients with gout is said to increase lysosomal permeability with resultant cell injury. Some photosensitivity reactions are due to lysosomal membrane damage when certain pigments, e.g. porphyrin, taken up by lysosomes, release energy on exposure to light of appropriate wavelength. Cortisol and chloroquine, drugs known to stabilise lysosomal membranes, diminish cell damage in vitamin A poisoning and some photosensitivity reactions. In many



Fig. 2.15 Electron micrographs of human erythroblasts. *Above*, mitochondria of normal erythroblast. *Below*, mitochondrion of erythroblast from a patient with sideroblastic anaemia. The mitochondrion is swollen and has an electron-lucent matrix in which lie electron-dense masses of iron-containing material. $\times 37\,400$. (Dr. A. M. Mackay.)

forms of cellular injury, however, the 'suicidal' release of lysosomal enzymes into the cytoplasm does not seem to be an important factor. For example, autolysis by lysosomal enzymes in cells injured by hypoxia, carbon tetrachloride and many other poisons occurs only after the cells have become necrotic.

Phagosomes containing membranous structures (e.g. damaged mitochondria) are commonly encountered in cells with focal cytoplasmic damage such as follows irradiation (Fig. 2.16), during starvation, in tumour cells, etc. The damaged parts of the cell are taken into an autophagic vacuole which coalesces with a primary lysosome to form a phagosome, and the activated hydrolases digest the contents of the vacuole. This process is accelerated by glucagon which acts via adenyl cyclase and cyclic AMP. Material resistant to digestion sometimes remains within a phagosome and forms one variety of **residual body** seen on electron microscopy. Lipofuscin pigment (p. 285) seems to originate from undigested lipid-rich material in this way. The hydrolysis of effete membranous structures and the re-utilisation of the products of digestion are probably of considerable importance in cellular economy.

Many inborn errors of lysosomal function have been described. Most of these involve deficiency of an enzyme and are characterised by progressive accumulation of the appropriate substrate within greatly swollen lysosomes. Some examples of the resulting 'storage diseases' are given on p. 28 and p. 32. Marked accumulation of metabolites in lysosomes with normal enzymes also occurs as a result of excessive production of substrate, e.g. α -chains of

haemoglobin in red cell precursors in β thalassaemia (p. 523) and lipids in macrophages in certain hyperlipoproteinaemias (p. 566). Storage of heavy metals, e.g. iron in the form of ferritin and haemosiderin and copper, occurs in lysosomes when cells are overloaded with these substances.



Fig. 2.16 Electron micrograph showing two autophagic vacuoles in adjacent cells of intestinal mucosa of mouse following radiation injury. Mitochondria and glycogen granules can be identified in the large, electron-dense vacuoles. $\times 57\,000$.

Abnormal storage of triglyceride fat

Triglycerides (or neutral fats) are glycerol esters of long-chain fatty acids and their storage in excess is a common and conspicuous feature of cell damage. Historically one of the first recognised features of sublethal cellular injury, its mode of development has been a centre of interest for many years. It is important to distinguish between an abnormal increase in the cells of adipose tissue (**pathological obesity**—or, when the condition is localised, **pathological**

adiposity) and the accumulation of fat in other types of cell (**fatty change**). The two processes are quite distinct although in pathological obesity fat also commonly accumulates in the liver cells.

Fat metabolism

Before outlining the main features of fat metabolism, it is important to appreciate that water-

insoluble lipids are transported in the plasma mainly as complexes of triglyceride fats, cholesterol and its esters, phospholipid and carrier protein (apoprotein). There are four major classes of such complexes in the plasma, namely chylomicrons and α , β and pre- β lipoproteins. These complexes differ in molecular size and in the proportion of their constituents, and so in density.

The apoproteins are essential for the aqueous solubility of the lipoproteins: the three main ones are called *apoprotein* (or *apo-*) A, B and C, and are synthesised mainly in the liver.

Over 95 per cent of the triglyceride in the diet is normally absorbed in the small intestine: much of it is hydrolysed in the gut into free fatty acids and monoglycerides, but these are re-esterified in the jejunal mucosal cells, and incorporated into large (up to 100 nm diam.) particles, the **chylomicrons**, which contain apo-B and -C. The chylomicrons pass via the gut lymphatics to the plasma, and are mainly responsible for its turbidity after a fatty meal. They consist of 90 per cent triglyceride fats, which are carried in the plasma mainly in this form. Reduction in size of the chylomicrons, and hydrolysis of fat, is effected by lipoprotein-lipase in the plasma, and the resulting glycerol and fatty acids are taken up by the cells of various tissues, including the liver. Some of the fatty acid is oxidised to provide energy, but much of it is re-esterified to triglyceride which is incorporated into **pre- β or very low density lipoproteins**: these are then secreted into the plasma and provide a means of transporting triglycerides in water-soluble form. The pre- β -lipoproteins, which make use of apo-B and apo-C, re-cycle through the liver, but are also taken up by the cells of the fat depots and other tissues, where the triglyceride is either used for energy or stored.

Triglyceride stored in adipose tissue is continuously being hydrolysed, and the fatty acids are secreted into the plasma where they are complexed with albumin. These so-called **free fatty acids** are taken up by muscle and other cells of high metabolic activity for the production of energy: that taken up by the liver cells is either oxidised or re-esterified to form triglyceride.

The α or **high density lipoproteins** of the plasma are assembled in the liver, and make use of apoprotein A. Their function is uncertain

but they seem to be of importance in the transport of cholesterol (see α -lipoprotein deficiency, p. 29). The β or **low density lipoproteins**, which contain apoprotein-B, transport cholesterol from the various tissues which synthesise it (especially the intestinal mucosa) to the liver, where it is added to that synthesised by the liver cells: much of it is then secreted into the bile.

The control of these major metabolic processes is influenced by food intake and also by various hormonal and emotional factors: insulin stimulates the deposition of triglyceride in the adipose tissue depots; adrenal hormones (probably catecholamines and corticosteroids acting together) stimulate hydrolysis in the depots and release of fatty acids into the plasma, as does growth hormone and also thyroxine. The rate of uptake of triglyceride and fatty acids by various other tissues, particularly the liver, is dependent on their concentrations in the plasma. Starvation results in release of fatty acids from the depots, and this is suppressed after a fatty meal. In addition to that provided by the diet, triglyceride is synthesised within the body, particularly in the liver and adipose tissues, from glucose, amino acids and fatty acids, and enters the metabolic pathways outlined above.

Fatty change

This is the accumulation of fat in cells other than adipose tissue cells, and was previously subdivided into fatty infiltration and fatty degeneration. It is due to imbalance between fat and fatty acids entering the cell and the rate of utilisation or release of fat by the cell. Probably all parenchymal cells which accumulate an abnormal amount of fat are injured. Even the gross fatty change seen in the liver in obesity (p. 667) can be regarded as a form of injury to the liver cells due to the disturbed fat metabolism resulting from overeating.

Because of its major role in fat metabolism, the liver requires special consideration in fatty change. However, fatty change is seen not only in the liver cells, which are usually most seriously affected, but also in various other organs and tissues. The cells most prone to undergo fatty change are the parenchymal cells of the various organs, and skeletal and heart muscle cells, i.e. the cells which because of their special-

ised functions have a high metabolic activity. Part of their energy is normally supplied by the oxidation of fatty acids, provided largely by uptake from the plasma of free fatty acids released from the fat depots, and lipoproteins secreted mainly by the liver.

Microscopic appearances

In fatty change of most organs, small droplets, consisting mainly of triglyceride, appear in the cytoplasm of the affected cells. Even in an advanced stage the droplets remain small and discrete, and do not greatly enlarge the cell (Fig. 2.17). In the liver, however, they may fuse to form much larger droplets (Fig. 2.18) and the liver cells may be greatly distended. In most instances, the centrilobular cells* are affected first and most severely, but in phosphorus poisoning the change may be very severe and yet confined to the cells in the outer part of the lobules. The distribution of these lesions has not been fully explained.

Electron microscopy shows the fat globules to lie free in the cytoplasmic matrix, without a limiting membrane.



Fig. 2.17a Fatty change of heart muscle, stained with osmic acid. Note the very numerous minute intracellular droplets arranged in rows. $\times 380$.

Causes of fatty change

The four major causes of fatty change are (a) hypoxia, (b) starvation and wasting diseases, (c) metabolic disorders, and (d) numerous chemicals and bacterial toxins.

(a) **Hypoxia.** The hypoxia resulting from chronic anaemia is a common cause of fatty change in the various organs and tissues. In the liver it occurs especially in the central zone of lobules, which receives the poorest supply of oxygenated blood, while in the heart it is more marked in parts farthest from the arterioles, that is, it is pararterial in distribution. In the heart, the change may be seen through the endocardium as a fine mottled pallor ('thrush breast') of the myocardium of the left ventricle and papillary muscles.

The fatty change commonly present in the cells of rapidly growing tumours is probably of the same nature, although in this instance there is inadequate blood supply or flow, and not just hypoxia.

(b) **Starvation and wasting diseases.** Fatty change is observed in the liver and to a smaller extent in the myocardium and elsewhere in



Fig. 2.17b Fatty change of tubules of kidney, stained with osmic acid. $\times 325$.

* The lobule, centred on a hepatic venule, is still widely used, as above, to describe the location of changes in the hepatic parenchyma. However, it now seems more logical to consider the functional unit as the acinus centred on a portal tract, as explained on p. 661.



Fig. 2.18 Fatty change of liver. The cells are filled with large globules of fat. (Stained with Sudan IV.) $\times 190$.

some patients who, previously well nourished, have died of a wasting disease such as gastric carcinoma or pulmonary tuberculosis. In other instances of equally or even more severe wasting, fatty change is not found. This was explained by Dible and his co-workers in 1941 who demonstrated experimentally that fat accumulates in the cells of the liver and other tissues under conditions of near-starvation so long as some adipose tissue remains. Once the fat depots are depleted, the fatty change disappears. The low food intake in wasting diseases leads to excessive lipolysis in the depots and release of fatty acids into the blood: these are taken up in increased amounts by the cells of various tissues and converted to triglyceride. Failure of carbohydrate metabolism resulting from the low caloric intake may also be of importance by interfering with intracellular oxidative breakdown of fatty acids, particularly in the liver.

Where a wasting disease is attributable to a toxic condition or complicated by severe infection, this may further impair hepatic fat metabolism, as explained below: a good example is provided by infantile gastro-enteritis due to

certain strains of *Esch. coli*, in which dietary intake is severely impaired by anorexia, vomiting and diarrhoea, and the liver is subjected to toxins absorbed from the infected gut. The liver usually shows gross fatty change in fatal cases.

(c) **Metabolic disorders.** The fatty change in uncontrolled **diabetes mellitus** is attributable mainly to excessive release of fatty acids from the fat depots and impaired carbohydrate metabolism, both of which are a consequence of deficiency of insulin. The situation is thus similar to that in starvation, and in both conditions oxidation of fatty acids in the liver is incomplete and partial breakdown products, acetoacetic and hydroxybutyric acids, escape into the blood, resulting in ketosis. The subject is discussed more fully on p. 1031.

Fatty change is common in the liver in various inborn errors of carbohydrate metabolism including some types of glycogen storage disease (p. 30), galactosaemia, and also in disorders of amino-acid metabolism.

(d) **Chemical and bacterial toxins.** Of the many simple chemicals which can cause fatty change, phosphorus, carbon tetrachloride and puromycin are well-known examples. As in fatty change from other causes, the liver is usually most severely affected, but the changes are widespread, and may involve not only parenchymal cells, but also vascular endothelium and connective tissue cells. Fatty change is also a feature of severe infections, e.g. typhoid, smallpox and septicaemias.

Two factors are involved in the production of fatty change by chemicals and toxins. Firstly, they directly injure the cells; secondly, they produce anorexia and often vomiting, and the low calorie intake results, as described above, in increased mobilisation of fatty acids from the depots. The nature of the cell injuries by many chemicals and toxins and the way these cause fatty change is by no means fully understood.

As a result of recent experimental investigations, it has been established that fatty change in the liver is attributable largely to reduced production of lipoprotein. The triglyceride which would normally have been released as lipoprotein thus accumulates in the liver cells. This applies to fatty change induced in rats by administration of chlorinated hydrocarbons, ethionine, phosphorus, puromycin, orotic acid

or a diet deficient in choline, each of which interfere with lipoprotein synthesis in different ways. There is some evidence that choline deficiency impairs the production of phospholipids, which are an essential constituent of lipoproteins.

It still remains unexplained how the various chemicals and toxins which cause fatty change in the liver affect also the cells of various other tissues. With the exception of the intestinal mucosa, which shares with the liver the property of converting fat to lipoprotein, tissue cells in general expend fat mainly by oxidative breakdown. It therefore seems likely that the various chemicals and toxins which produce widespread fatty change interfere in some way with this latter process.

Effects of fatty change

Fatty change results from cell injury, but varies in degree in different types of cell injury. For example, liver cell necrosis in viral hepatitis is not preceded or accompanied by any significant degree of fatty change, but there is severe fatty change associated with liver cell necrosis in phosphorus poisoning. Also the gross fatty change in the liver which may accompany obesity or alcoholism is not itself usually accompanied by severely impaired hepatic function.

Fat accumulates less rapidly in the other organs than in the liver but, as in the liver, the important factor is the severity of the cell injury, which is not reflected by the degree of the fatty change.

Fatty change in the heart may indicate severe myocardial injury, from which heart failure may result. For example, in severe anaemia attributable to a lesion requiring surgical intervention, such as recurrent haemorrhage from a peptic ulcer, it is important that, when practicable, the anaemia should be treated and time allowed for the myocardium to return to normal before any major operation is undertaken. The administration of a large volume of blood or packed red cells over a short period carries a risk of overloading the impaired myocardium, especially if followed immediately by major surgery.

Pathological obesity

Obesity, the accumulation of excessive amounts of adipose tissue, is a subject in which it is dif-

ficult, if not impossible, to draw a sharp dividing line between physiological and pathological states. However, there is no doubt that gross obesity is harmful, and must be regarded as pathological.

Causes. Basically, obesity is very simply explained, being due to a dietary intake of calories in excess of those expended to provide energy for the body's metabolism. It is thus attributable to overeating, particularly of carbohydrates and fats, often combined with lack of exercise. Attempts to demonstrate metabolic differences between fat and thin people, e.g. in efficiency of intestinal absorption or in basal metabolic rate have, in general, been unsuccessful and the main problem of obesity appears to be the cause of overeating, a subject involving psychological factors which will not be discussed here. Some individuals, however, seem to be predisposed to obesity more than others, and genetic factors may be involved, as in some inbred strains of animals. It has been observed that when healthy young adults are given a high calorie diet and kept at rest in bed, those who are overweight gain more weight than the thinner subjects. The activity of the thyroid gland, by influencing the rate of general metabolism, has an important influence on energy expenditure, and the pituitary, adrenals and gonads all influence the amount of fat deposited. Damage to the hypothalamus with deficiency of pituitary secretion in early life leads to adiposity along with failure of sexual development, and there is evidence that some forms of obesity in the adult are of similar causation. Extreme degrees of adiposity can be induced in rats by small precisely placed experimental lesions in the tuber cinereum, the mode of action of which appears to be the development of a voracious appetite.

Gross abnormalities of the hypothalamus or of endocrine function have not been demonstrated in the great majority of obese subjects investigated, but it may be that more subtle variations in the functioning of these organs are of importance.

Structural changes. Apart from the increase in size of the normal depots, e.g. the subcutaneous tissue, the omentum, retroperitoneal tissues and epicardium, adipose tissue in obesity may extend to sites where it is normally absent. For example, in pathological adiposity of the heart, adipose tissue extends along the lines of con-

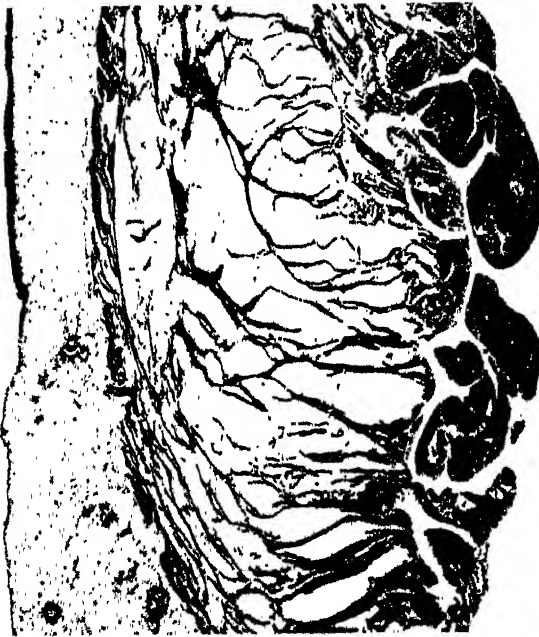


Fig. 2.19 Pathological adiposity of heart. The whole thickness of the wall of the right ventricle is infiltrated with adipose tissue extending from the epicardial layer between the muscle fibres which are

consequently atrophied. Even the columnae carneae are involved. $\times 6$.

nective tissue through the heart wall (Fig. 2.19), and leads to atrophy of the muscle fibres and consequent weakening of the wall to such an extent that the right ventricle may rupture. A similar accumulation of stromal fat is seen occasionally in the pancreas.

In obese individuals, the liver may be grossly enlarged by accumulation of large droplets of fat in the liver cells. This is discussed on p. 667.

Effects. Apart from the limitations imposed on physical activity by obesity, it has long been recognised, and notably by life insurance companies, that obesity is associated with a reduced expectation of life attributable to an increased incidence of high blood pressure, coronary artery disease, heart failure, chronic bronchitis and respiratory infections, and late-onset diabetes.

The high and increasing incidence of obesity in affluent societies poses a major health problem.

Abnormal storage of other lipids

The lipids of the body other than triglyceride are chemically very heterogeneous and include sterols (cholesterol and its esters), phospholipids and complex lipids (e.g. glycolipids). These substances are frequently united with proteins to form lipoproteins, some of which constitute the insoluble membranes of cells, while others are soluble and play an important part in the transport of triglyceride in suspension in the plasma. Many diseases are known in which one or more of these substances accumulate in cells in abnormal amounts. By far the most important is **atheroma**, a poorly understood disorder in which various lipids, including sterols, phospholipids and triglyceride, accumulate in the intima of arteries and cause narrowing or occlusion of the lumen with consequent impairment of blood flow.

Most of the other disorders are much less common but illustrate interesting principles. Pathological accumulation of lipids other than triglycerides within cells may develop in the following ways.

Inherited deficiency of lysosomal enzymes

Cells may accumulate lipids derived from the normal turnover of the membrane material bounding the cell surface and organelles. This occurs in individuals lacking an enzyme necessary for the catabolism of membrane-derived lipids, and the accumulation of lipid is often especially prominent in macrophages because of their phagocytic function. A good example is Gaucher's disease in which there is a genetic deficiency of a lysosomal β -glucosidase. Normally, old worn-out red cells are phagocytosed and digested by macrophages, mainly in the spleen. In Gaucher's disease the glucocerebrosides of the red cell membrane accumulate in phagosomes in the macrophages, which consequently become greatly enlarged and develop a characteristic appearance (Fig. 18.3, p. 567) and are termed Gaucher cells. Their accumulation causes gross splenomegaly, hepatomegaly, and anaemia due to replacement of the haemo-

poietic tissue by Gaucher cells. Another form of Gaucher's disease affects the cerebral neurones (p. 567).

In marked contrast to Gaucher's disease, the lesions of Krabbe's disease are confined to the central and peripheral nervous systems because galactocerebroside, the substrate of the deficient enzyme galactocerebroside β -galactosidase, is a major component only of myelin. This recessive lysosomal enzyme defect causes, in infancy, severe central and peripheral nervous dysfunction due to unexplained demyelination of axons. Krabbe's disease is also known as 'globoid leukodystrophy' since the white matter of the brain is particularly affected and the lesions contain macrophages: these are globoid in shape because they are distended with PAS-positive galactocerebroside which has characteristic electron-microscopic appearances. Gaucher's disease and Krabbe's disease illustrate the great differences between lesions due to inherited deficiency of enzymes responsible for intracellular breakdown of lipids and demonstrate some of the reasons for these differences.

Disordered plasma lipid transport

Several inherited abnormalities of plasma lipoproteins are known which lead to massive accumulation of lipid in macrophages. These cells become spherical and enlarged due to the presence of numerous lipid vacuoles which give the cytoplasm a foamy appearance (Fig. 2.20). The predominant lipids stored are cholesterol esters

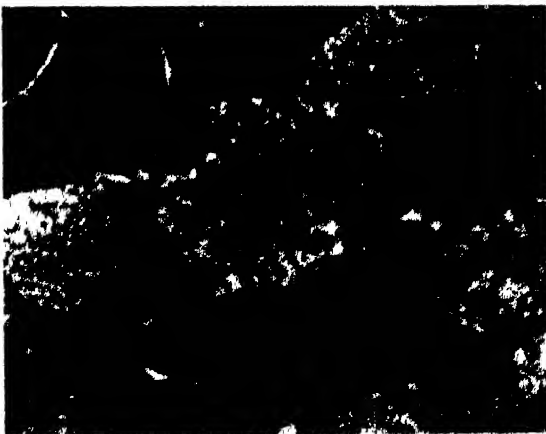


Fig. 2.20 Macrophages distended with multiple fine droplets of doubly-refractile lipid. Photographed through crossed polarising prisms. $\times 500$.

which are seen to be doubly refractile when viewed under the microscope with polarised light (Fig. 2.21). Large accumulations of these cells (often accompanied by extracellular lipid deposits) may form nodules, called **xanthomas** because of their yellow and tumour-like appearance.



Fig. 2.21 Deposition of doubly-refractile lipid in early atheroma of aorta, photographed by polarised light.

The rare **familial α -lipoprotein deficiency** (Tangier disease) is due to absence of apoprotein A. There is a moderate increase in the plasma level of triglyceride which is associated with β -lipoproteins as unstable complexes. These are phagocytosed by macrophages in the spleen, liver, tonsils and lymph nodes, all of which organs become enlarged due to accumulation of cholesterol esters, presumably derived from the β -lipoprotein. Great enlargement of the tonsils, orange in colour due to the stored lipid, is a unique finding in this disease.

Familial hyperchylomicronaemia, apparently due to inherited deficiency of serum lipoprotein lipase, is characterised by a milky appearance of the serum, even after overnight fasting, due to chylomicrons which are not cleared from the plasma at the normal rate. There is lipid storage in macrophages in the enlarged liver and spleen, and xanthomas may be found in the dermis. It is of interest that in both of the above diseases the stored lipid is predominantly cholesterol ester, although the main elevation in serum lipid affects the triglycerides in the chylomicrons. Presumably the latter are broken down following phagocytosis by the macrophages, leaving a steadily increasing residue of less digestible esters of cholesterol. The steroid nucleus cannot readily be broken down and is mainly disposed of by excretion.

Familial hyperbetalipoproteinaemia is a relatively common disorder, inherited as a Mendelian dominant, in which the plasma β -lipoprotein is greatly increased. Cholesterol, an important integral part of the lipoprotein molecule, is correspondingly raised, hence the old name 'familial hypercholesterolaemia'. There is little elevation in triglycerides and the plasma is not milky. Atheroma, fatal even in childhood in rare homozygous individuals, and in many cases xanthomas affecting skin and tendon sheaths, increase with time and parallel in severity the plasma lipoprotein abnormality.

Many kinds of cells, including fibroblasts and aortic smooth muscle cells (but not liver or intestinal epithelium), have surface receptors for the apoprotein of β -lipoproteins. High lipoprotein concentrations in plasma and interstitial fluid saturate these receptors and this leads to diminished production of the rate-limiting enzyme for cholesterol synthesis in these cells (i.e. there is a negative feed-back control). In individuals homozygous for hyperbetalipoproteinaemia, the receptors are absent and cholesterol synthesis is greatly increased. Heterozygotes have half the normal number of receptor sites and this is compensated by raised levels of β -lipoproteins.

These and other inherited lipoprotein disorders illustrate the varied metabolic abnormalities underlying excessive storage of cholesterol esters in foam cells in different parts of the body. Hyperlipoproteinaemia also develops secondary to other diseases; for example, transitory

hyperchylomicronaemia is encountered sometimes when severe diabetes mellitus is inadequately controlled and secondary hyperbetalipoproteinaemia, reversible by thyroxine therapy, is seen in hypothyroidism. These and other disorders (e.g. obstructive jaundice, pancreatitis, alcoholism and nephrotic syndrome) may be associated with various forms of hyperlipoproteinaemia.

Other lipid depositions

Submucosal aggregates of foam cells are often found in the gallbladder giving it a 'strawberry' appearance (Fig. 20.64, p. 710); presumably this results from intracellular storage of part of the cholesterol that is normally reabsorbed from the bile.

The important subject of atheroma will be considered later (p. 362). Suffice it to say here that pressure filtration of lipoprotein from the plasma into the intimal layer of the arteries leads to a difficult problem in disposing of the associated cholesterol by the local population of modified smooth muscle cells. At first the filtered lipid is found within these cells but these are quickly overwhelmed and most of the accumulated lipid eventually lies in an extracellular position. As already indicated, atheroma is particularly prone to develop in individuals with certain of the inherited lipoprotein abnormalities; it seems to result also from modern western dietary habits which lead to alterations in serum lipoproteins and lipids.

Abnormal storage of glycogen

The normal human body contains approximately 500 grams of glycogen, present mainly in muscle and liver cells, but also found in small amounts in the other tissues. Glycogen is a water-soluble branched polymer, composed exclusively of glucose units, and is broken down by enzymes (glycogenolysis) to provide glucose needed to meet increased energy requirements, e.g. during muscular exercise. The depolymerisation is effected mainly by phosphorylase enzymes which liberate glucose 1-phosphate, and this in turn is converted into glucose 6-phosphate which can be used for the

intrinsic metabolic needs of the cell. For the maintenance of blood glucose levels during fasting and exercise, glucose 6-phosphate must be converted to glucose before release from the cell, and this happens almost exclusively in the liver, but also in the kidney, both of which contain the necessary enzyme, glucose 6-phosphatase.

Largely as a result of the brilliant biochemical studies of G. E. Cori, a number of distinct inherited **glycogen storage diseases** have been recognised, each associated with a different single enzyme deficiency. In several forms of

the disease, the affected organs are enlarged due to an increased content of glycogen. Histological examination of sections stained by haematoxylin and eosin reveals clear unstained material distending the affected cells (Fig. 2.22); histochemical confirmation of the nature of the material is given by Best's carmine, or the periodic acid Schiff (PAS) stain (with and without prior hydrolysis of the section with diastase), and this is most satisfactorily obtained with tissue fixed promptly in alcohol in which glycogen is insoluble. Glycogen also has a characteristic appearance on electron microscopy, occurring as dense particles larger than

ribosomes; these sometimes form rosette-like clusters (Fig. 2.23).

In *von Gierke's disease* (Cori Type 1) there is deficiency of glucose 6-phosphatase, an enzyme which is associated with the endoplasmic reticulum. The resulting failure to convert glucose 6-phosphate to glucose for release into the circulation leads to hypoglycaemia and a tendency to increased glycogen storage in liver and kidney, the organs which normally contain glucose 6-phosphatase. As a consequence of the excessive glycogen storage, these organs become enormously enlarged. The disordered carbohydrate metabolism leads to increased lipogenesis and raised serum levels of triglyceride (as the very low density pre- β -lipoprotein), xanthomatosis, obesity, fatty change in the liver and ketoacidosis. The affected children also show retarded growth.

Pompe's disease (Cori Type 2) has recently been shown to result from an inherited deficiency of the lysosomal enzyme acid α -glucosidase which is normally present in all tissues and presumably hydrolyses the small amount of glycogen present in phagosomes as a result of



Fig. 2.22 Myocardium in Pompe's disease *a*, compared with normal myocardium. *b*. The affected muscle fibres are distended with glycogen and appear vacuolated. $\times 460$.

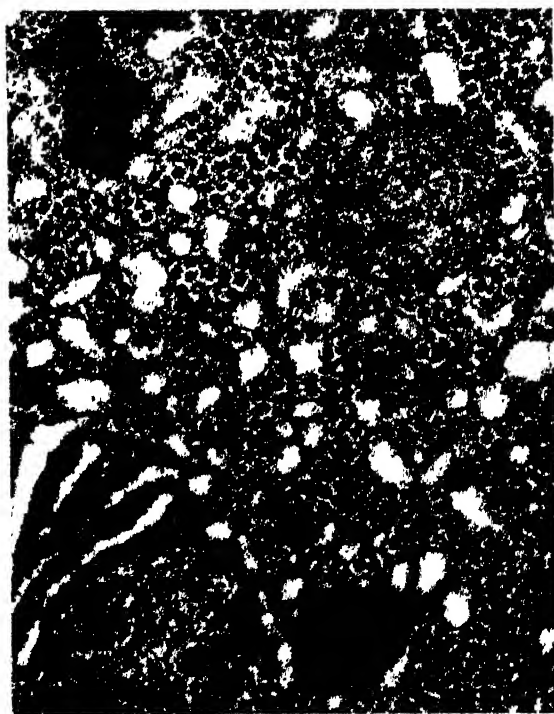


Fig. 2.23 Electron micrograph of normal liver showing the characteristic small, dense, round particles of glycogen. The large dense organelle at the top of the field is a residual body. $\times 30\,000$.

autophagy. In Pompe's disease, the glycogen in autophagosomes persists and accumulates, being inaccessible to the general cytoplasm with its normal complement of phosphorylase and other enzymes in the major glycogenolytic pathway. Accordingly much of the stored glycogen is seen by electron microscopy to be within greatly enlarged lysosomes, and it seems likely that the serious cellular dysfunction encountered in Pompe's disease is the result of lysosomal rupture and spilling of harmful hydrolases into the general cytoplasm of the cell. This explains most of the features of Pompe's disease, namely generalised glycogen storage—e.g. in myocardium, skeletal muscle, nervous and lymphoid tissue, and peripheral blood leukocytes—enlargement of organs, cardiac failure, mental deficiency, severe muscle weakness and absence of hypoglycaemia.

Other extremely rare forms of glycogen storage disease include deficiency of muscle phosphorylase (**McArdle's syndrome**) which results in abnormal accumulation of glycogen in skeletal muscle and rapid muscle fatigue without hypoglycaemia, and deficiency of liver phosphorylase which causes hepatomegaly and hypoglycaemia.

Another interesting disorder is the deficiency

of the enzyme responsible for the branching of the tree-like glycogen molecule during its synthesis: in this condition the stored glycogen has an abnormal fibrillary structure seen by electron microscopy and is resistant to breakdown by phosphorylase.

In contrast to the rare disorders described above in which deficiency of a particular enzyme satisfactorily explains most of the observed pathological findings, increased glycogen storage is found in many other disorders but the mechanisms involved are usually obscure. In diabetes mellitus utilisation of glucose is seriously impaired and excess of glycogen is deposited in cardiac muscle, in Henle's tubules in the kidney, in liver cell nuclei, in polymorphonuclear leukocytes and in the hydropic β -cells of the pancreatic islets seen in early Type I diabetes (p. 1032). The glycogen content of skeletal muscle is, however, reduced. More glycogen than normal is found in polymorphs in acute inflammation, in recently formed pus and in the peripheral blood when there is a leukocytosis. Glycogen is abundant in embryonic tissues and in some malignant tumours arising from cells normally rich in glycogen, e.g. liver and renal tubular epithelium. It is particularly constant in clear cell carcinoma of kidney.

Cell damage due to ionising radiation

Because of the widespread use of radiation and radioactive materials in industry and medicine the study of their effects on living matter is now of great importance. Of all the branches of radiation biology—molecular, sub-cellular, cellular, organ and whole animal—cellular radiation biology is the most instructive in the present state of knowledge. This is firstly because the measurement of its effects on proliferation of cells in culture makes quantitation fairly easy. Secondly, it is mainly by the study and analysis of cellular effects that information can be obtained on molecular and sub-cellular processes in the development and repair of radiation damage. Thirdly, it appears that effects on whole organs or whole animals can often, to a first approximation at least, be described or explained on the basis of cellular injury. This third principle will be illustrated below in terms of the impaired

ability of sub-lethally irradiated cells to divide, although it must be emphasised that the effects of radiation on the individual cannot easily be explained fully by its effects on his component cells.

Radiation causes its effects by transferring energy to the substance through which it passes. This energy can produce two changes, excitation or ionisation. Excitation is a change in the energy state of some electrons or charged parts of molecules. Radiations which, like ultra-violet light, produce excitations only, are of very low penetrating power and are not discussed in this section. Other radiations produce mixtures of ionisations and excitations.

Measurement of radiation

No widely useful and accurate biological method of measuring radiation dose has so far

been developed. This is because the ultimate biological effects produced are very complex. The standard methods of measuring radiation dose are purely physical. The first is based on the application of a voltage to an air-filled chamber through which the ionising radiation passes. This voltage has the effect of separating the positive and negative electric charges produced by the ionisations. The flow of these charges under the influence of the applied voltage constitutes a small electric current which can be measured with great sensitivity and accuracy. The unit of *exposure* measured by this method is thus based on the electric charge per mass of air (coulomb/kilogram) and was called the *röntgen* (R). This method ignores the excitations produced by the radiation, but *exposure* expressed in R agrees roughly with the absorbed dose (see below) measured in rads.

The second method of measuring radiation dose is based on the measurement of the total energy absorbed from a beam of radiation by a solid material. The material must be one in which no radiation-produced chemical energy can be stored, all of it being transformed into heat which can be measured by sensitive calorimetry. The unit of absorbed dose is the *gray* (Gy), called after the British scientist who did fundamental work on the oxygen effect described below. Absorbed dose used to be expressed in units called *rads*. The new unit, Gy, equals 100 rads.

Much research has been done on the effects of radiation on aqueous solutions; these are mediated primarily through the decomposition products of the water, and probably reflect the initial damage produced in living biological material. However, this initial damage has never been directly observed. The effects which are observed are the result of an interaction, between the disruptive effects on the complex of normal biochemical processes arising from the initial molecular damage, and the response of the cell or organism in trying to overcome the disruption. From an analysis of these effects much has been learned about the nature of the disruption and repair, but very much more still remains unknown.

Types of ionising radiation

Ionising radiations fall into three categories; electromagnetic radiation, charged particles

and uncharged particles. The ionisation is caused by the charged particles—electrons, protons or heavy nuclei: electromagnetic radiation and uncharged particles produce charged particles.

All of these radiations can be administered to animals and human beings externally or internally. External sources include x-ray machines, electron or other charged particle accelerators, high activity radiation γ -ray sources, neutron generators, nuclear reactors and atomic bombs. Internal irradiation arises from the ingestion of any of the hundreds of known radionuclides (radioactive isotopes) many of which are used for medical diagnostic or therapeutic purposes, for commercial non-destructive testing and for irradiating materials for industrial purposes; they are produced in nuclear reactors. The dosage has, in the past, been expressed as *curies* (Ci), but the modern unit is the *becquerel* (Bq) after the French scientist who discovered radioactivity. One Ci equals 3.7×10^{10} Bq. The best known severe radiation damage from accidental ingestion of radioactive material is that produced by radium which was once used extensively in luminising paints. Some sufferers from radium poisoning have been under medical supervision for up to fifty years and the effects are well documented. Today, in spite of strict regulations and control, occasional accidents occur in nuclear reactors, industry and hospitals, which result in significant radiation of personnel. Ionising radiation for diagnosis and therapy, however, has a secure and important place in modern medicine and has been essential for many research purposes.

Cellular radiation effects

Measurement of cellular radiation effects is based on the survival curve, i.e. on the ability of cells to multiply in appropriate environmental conditions. One of the most marked effects of radiation is the destruction of this ability. The number of cells in which the ability survives can be readily measured by counting the number of clones in cultures of the irradiated cells.

The percentage 'survival', that is, the percentage of cells which still retain the ability to produce clones after irradiation, is plotted on a logarithmic scale and the radiation dose is plotted on a linear scale (Fig. 2.24). Since log

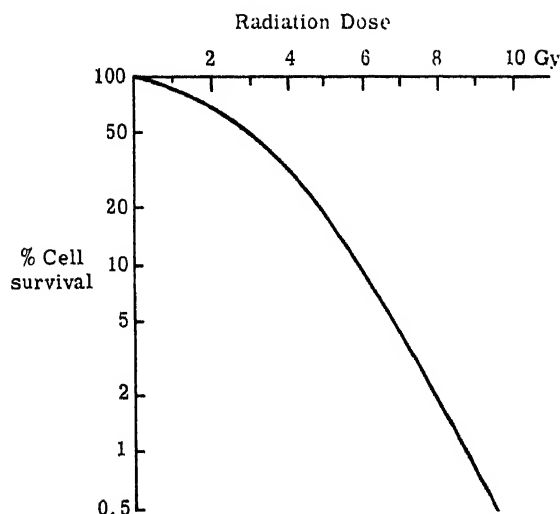


Fig. 2.24 Radiation injury, showing the relationship between percentage cell survival and dose of ionising radiation.

percentage survival decreases continuously with increasing dosage, the origin is placed in the top left corner of the figure. The resultant curve usually approximates to a straight line after an initial shoulder. This shoulder shows that an accumulation of sub-lethal damage is necessary before an observable effect is produced. After sufficient dose has been given to reach the straight portion of the survival curve, no further accumulation of sub-lethal damage occurs. If, however, the cells are allowed to recover for some hours, it is found that once again a considerable accumulation of sub-lethal damage is necessary before the radiation effects are produced with maximum efficiency. It appears that the shoulder therefore represents the repair process by which the cells can overcome some of the damage.

The slope of the portion of the survival curve which is almost straight on the log-linear plot of Fig. 2.24—the exponential part—represents the rate at which additional lethal damage is produced as additional dose is administered. The fact that this portion of the curve is almost straight means that, irrespective of the damage already created, a given amount of irradiation always reduces the survival by the same fraction. This kind of relationship, in which increasing dosage decreases proportionately the fraction of surviving cells, is indicative of an effect controlled by probability. In this case it is usually regarded as the probability of the track of ionisations, produced by charged particles,

causing damage to small discrete targets in individual cells.

The available evidence suggests that, for each type of radiation, no great differences exist in the slopes of survival curves for various types of mammalian cells, unless they are hypoxic when irradiated. The magnitude of the shoulder, however, is dependent on the history, the environment, and the biochemical condition of the cells.

Linear energy transfer effect. Survival curves referring to different types of radiation have different shapes. The reason for this is that although, for equal doses of different radiations, the total number of ionisations are equal, the geometrical distribution of these ionisations in the cell can vary greatly. These large variations in distribution have a considerable effect on the mean number of ionisations necessary to produce the kind of biological damage which, after complex development, results in a cell losing its ability to continue to divide.

Heavy charged particles moving relatively slowly produce dense columns of ionisations, while electrons produce sparse lines of ionisations with occasional small clumps (Fig. 2.25).

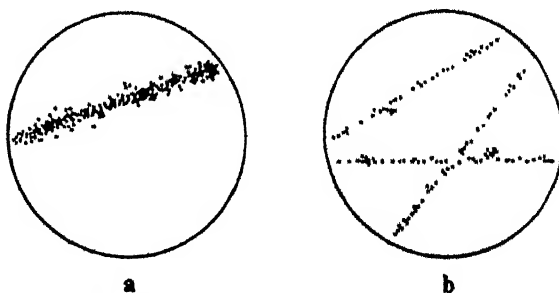


Fig. 2.25 The distribution of ionisations for high LET radiation (a) and low LET radiation (b).

The average amount of energy deposited per micron of track of a particle or photon is called the linear energy transfer (LET) and can be used to characterise the quality of the radiation. A number of ionisations grouped closely together are more likely to initiate a lethal chain of biological events than the same number of ionisations widely separated. The dense group of ionisations also prevents the cellular repair mechanisms from acting effectively. Thus the survival curves for high LET radiation such as α -particles show no shoulder, and have a steeper slope than those for low LET radiation such as x-rays or electrons (Fig.

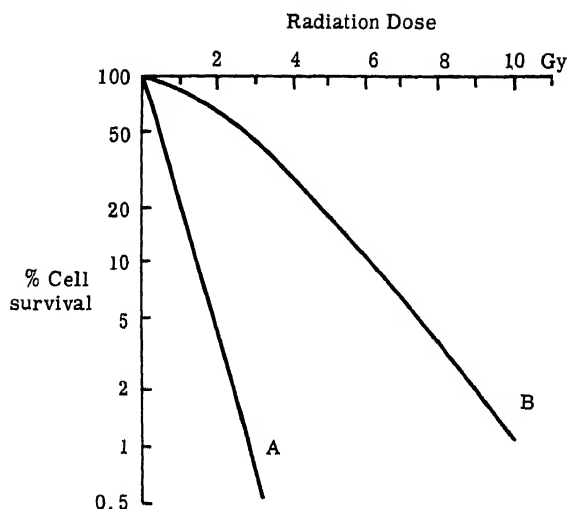


Fig. 2.26 The comparative cytotoxic effects of α -particles (A) and electrons (B).

2.26). For both these reasons a dose of high LET radiation produces far more biological damage than an equal dose of low LET radiation.

Effect of oxygen. Another factor which has a striking effect on survival curves is the presence or absence of oxygen. Oxygen has the ability to combine with freshly severed ends of molecular structures thus preventing them from rejoining, which they commonly do if the opportunity presents itself. Oxygen thus interferes with a natural recovery process—a different one from that which produces the shoulder—and causes a given dose of radiation to be much more damaging than it would be in hypoxic or anoxic conditions. The ratio of the doses required to reduce the survival to the same level in anoxic and normal conditions is called the **oxygen enhancement ratio (OER)** (Fig. 2.27).

There are many other radio-sensitisers and radio-protectors which affect different levels of recovery and repair processes. Estimation of the biological damage produced by a given dose of radiation must therefore take into account both the type of radiation, the environment in which it is administered and the time allowed for repair and recovery.

Tissue radiation effects

The effect of radiation in destroying the ability of cells to continue dividing has been discussed above in relation to individual cells. It is important because it plays a major part in causing

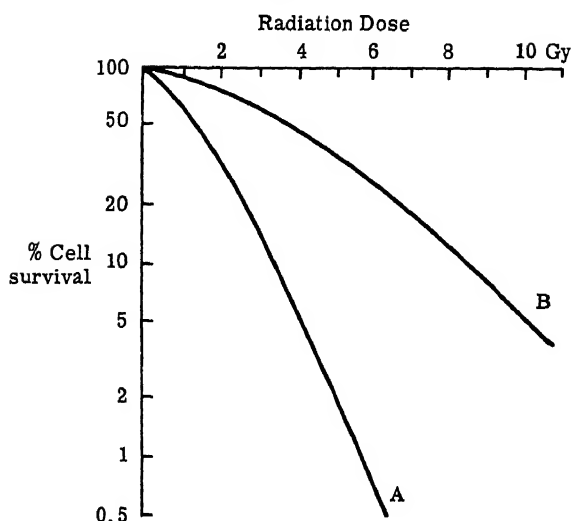


Fig. 2.27 The effect of oxygen on the cytotoxic effect of x-irradiation. A, high oxygen tension; B, low oxygen tension.

tissue effects. Cell function which is not related to mitosis and division is relatively insensitive to radiation. Much higher doses of radiation are required to produce gross changes in such function than are required to inhibit cell division. Hence the typical radiation effect on tissues arises from an inhibition of division.

Tissues whose cells are undergoing continuous controlled division, and whose integrity requires a continual flow of new cells, are therefore the first tissues to show the effects of radiation. The most obvious are the skin, the intestinal tract, the bone marrow and the immunity system. Similar considerations apply to the therapeutic use of radiation for malignant tumours, in which there is excessive division of cells. The initial effect upon a tissue is a reduction in cell numbers as the supply of new cells falls below the normal rate. The drop in cell numbers leads via homeostatic feedback mechanisms to a build up of the population of viable stem-cells from which new cells are produced, and to an increase in the rate of cell division. If this compensation is successful, then in due course enhanced production of cells not only restores the depleted population but commonly results in a temporary hyperplasia or overshoot before the cell numbers return to normal (Fig. 2.28). The inflammatory response to trauma follows injury by irradiation and commonly results in permanent structural changes, e.g. fibrosis (see p. 38).

The fall, rise, overshoot and return to

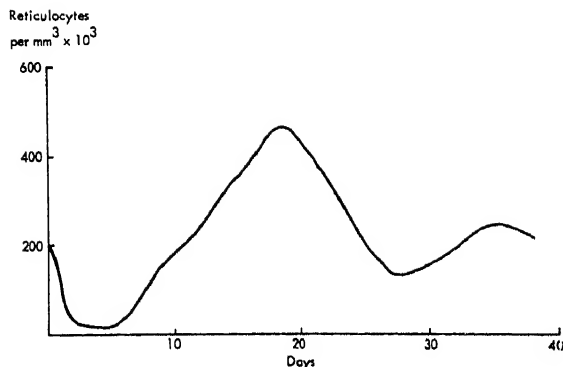


Fig. 2.28 Changes in the numbers of reticulocytes in the blood following 2 Gy whole-body irradiation.

normal of the cell populations exhibiting such behaviour after irradiation can be understood and explained in cellular terms if the appropriate homeostatic feedback mechanisms are known. The normal cell turnover controls the rate at which the cell population falls following irradiation. The extent of the fall depends on the percentage of surviving cells and the rate at which they can divide.

If the cell population of a tissue falls below a critical value the tissue can lose its functional effectiveness. In the cases of the intestinal tract and the bone marrow the result is death of the individual. A rapidly administered x-ray dose of about 8 Gy to the bone marrow and 12 Gy to the intestinal tract reduces the number of surviving cells to such low levels that the delay before an adequate production of cells can be re-established is long enough to allow the cell population to fall below the critical value.

The cells whose reproductive ability has been destroyed by the irradiation often remain in the tissue for some time. Their abortive attempts to divide or prepare to divide can produce gross abnormalities in cytological appearance (Fig. 2.29). Toxic products of cell disintegration can increase the damage. However, the damaged cells are no longer directly relevant to the course of events leading to permanent damage or repair. This course is determined by the number of surviving cells still capable of division and by the kinetics of proliferation in the tissue. Ultimately, however, if a large enough dose is given (about 18 Gy in a single exposure) a tissue condition described as the *limit of tolerance* is reached. Although the nature of this is not understood, it is the determining factor for

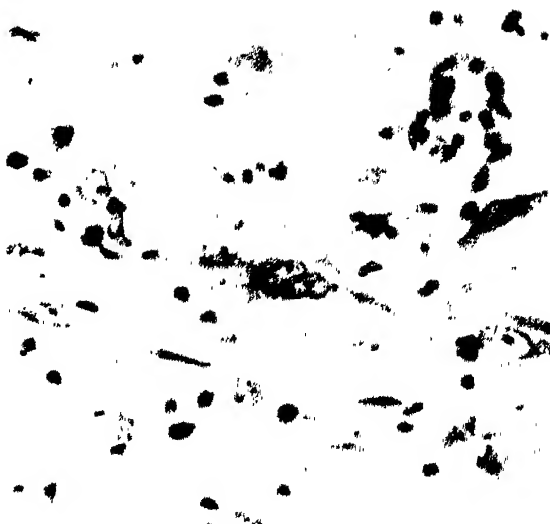


Fig. 2.29 Changes in the sub-epithelial connective tissue of the tongue following irradiation therapy for an epithelial tumour six years ago. Note the abnormally large connective tissue cells, one of which is binucleate. $\times 340$.

radiotherapy, and represents an accumulation of permanent, irreparable damage.

Tissues whose cells are long-lived and therefore are not normally dividing show very little effect after doses of several grays. Damage has been done, however, and becomes apparent if the cells are stimulated to divide, even after long intervals of time (Fig. 2.30). Examples of such tissue are the adult liver, adult thyroid and long-lived lymphocytes. Again the major effect is that many of the cells are unable to divide successfully when called upon to do so.

Radiation damage to the gonads may lead to infertility due to impairment of germ cell division. Errors also occur in copying the base sequence of DNA in the germ cells, the number of errors—mutations—being related to the radiation dosage as shown in Fig. 2.31. A given dose of radiation insufficient to cause infertility gives rise to the same total number of mutations irrespective of the number of individuals among whom it is distributed. As a corollary there is no 'safe' level of background radiation. It is known that radiation induces the development of malignant tumours, (p. 301) possibly by causing mutations in somatic cells (p. 319).

Microscopic appearances. Following a substantial dose of radiation there is a latent interval of hours or days before histological evidence of tissue injury is seen. As already explained, the



Fig. 2.30 Dividing lymphocyte in peripheral blood culture from a patient with bronchial carcinoma and spinal metastases, treated by five 5 Gy of ^{60}Co radiation to the lumbar spine. This cell shows the result of extensive chromosome breakage followed by random fusion of broken ends due to radiation. There are nine dicentric chromosomes, one possible tricentric, one acentric fragment and at least three other abnormal chromosomes. 44 centromeres can be counted, indicating elimination of two chromosomes. Aceto-orcein stain. $\times 2000$. (Professor M. A. Ferguson Smith.)

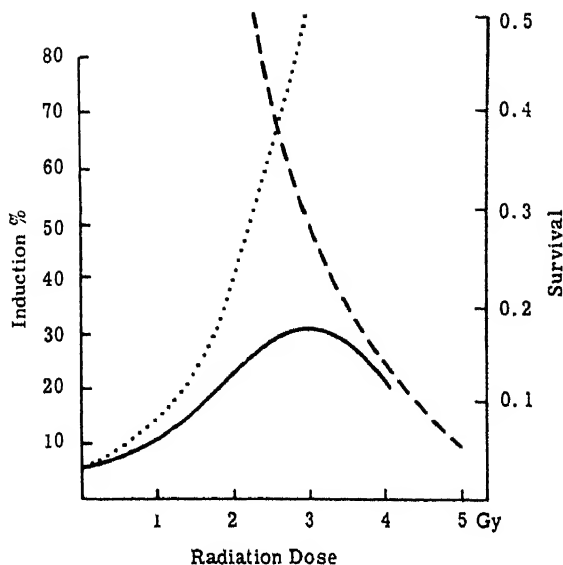


Fig. 2.31 The effect of dosage of radiation on mutation rate in mice. As the dose is increased, the mutation rate rises (dotted line), but the survival rate diminishes (interrupted line). The incidence of mutation-dependent abnormality, in this case leukaemia, is dependent on mutation and survival rates, and is shown by the continuous line.

damage depends on the dose and type of radiation, on the interval following exposure and on the tissue exposed. Early changes in the skin include dilatation of blood vessels and other signs of acute inflammation and these reflect acute tissue injury. With a single dose of 15 Gy mitotic activity of the basal cells is arrested, with subsequent loss of the epidermis and epilation. The walls of the dermal vessels are infiltrated with fibrin; later a characteristic concentric proliferation of intimal fibrous tissue is seen (endarteritis obliterans), followed by replacement with dense homogeneous (hyaline) collagen (Fig. 2.32). Large bizarre fibrocytic



Fig. 2.32 A small artery occluded by hyaline fibrous tissue following radiotherapy. (Same case as in Fig. 2.29.) $\times 126$.

nuclei are present in the dermal connective tissue (Fig. 2.29). With repeated exposure to radiation the dermal collagen becomes very dense and there is a tendency for the dermal fibrous tissue to become necrotic even years after exposure; persistent melanin pigmentation and vascular dilatation are also noted. Comparable changes found in other tissues following irradiation are described later in the appropriate chapters; the detailed findings depend, of course, upon the radio-sensitivity of the various types of tissue present, their turnover of cells and potential for cell multiplication, and the architectural features of the tissue.

Atrophy

By atrophy is meant diminution in size of a cell or reduction in the essential tissue of an organ due to decrease in the size or numbers of its specialised cells. Pathological atrophy has its prototype in the physiological atrophy of old age, which affects all the tissues, and notably the bones, lymphoid tissue and the sexual organs; and although some of the changes

occurring in old age are the result of atrophy of the gonads, this atrophy in its turn cannot be explained. The cause of **senile atrophy** is of course merely part of the larger question of what limits the duration of life. Atrophic specialised epithelial cells tend to lose their special features and to become de-differentiated, as may be seen in local atrophic changes

in the liver and kidneys. Senile atrophy is commonly accompanied by accumulation of the yellowish-brown pigment lipofuscin and the term *brown atrophy* is then applied. As already indicated (p. 23) lipofuscin represents indigestible lipid which forms residual bodies and is often the product of cellular autophagia.

An organ may be undersized as the result of imperfect development, and the term **hypoplasia** is then applied; for example, the hypoplasia of the genital glands which results from deficiency of the pituitary secretion in early life.

Causes of atrophy

1. Defective nutrition. This may be produced locally by arterial disease interfering with the blood supply to a part, when the reduction is not so severe as to cause necrosis. The functioning parenchymatous elements of the tissue then undergo atrophy, and sometimes there is also a concomitant overgrowth of fibrous tissue. This is often seen in the myocardium (Fig. 15.5, p. 402) and in the kidneys, in which small atrophic depressions result from narrowing of the lumina of the small arteries. **General atrophy** is seen in cases of starvation; emaciation depends chiefly upon utilisation of the fat of the adipose tissue but there is also a general wasting of the tissues. The various organs may thus diminish in weight, the liver and spleen are markedly affected, the kidneys and heart to a less though distinct degree, whilst the central nervous system is only slightly affected. In most cases of wasting disease, however, such as malignant tumours of the alimentary tract, chronic tuberculosis or suppuration, other ill-defined factors appear to contribute to the wasting. Various other forms of cellular damage and diminished cell production may thus come to be associated with atrophy; secondary anaemia, for example, is present, although slight or absent in wasting due to starvation alone. The term **cachexia** is often applied to a combination of wasting, anaemia and weakness.

2. Diminished functional activity. It is a general law that diminution in the catabolic processes leads to reduced anabolism and thus to diminution in the size of cells. When the function of a part is in abeyance the blood supply also diminishes. **Disuse atrophy**, as it is sometimes called, is seen when a gland, for example,

the pancreas, has its duct obstructed; its functional activity is thus stopped and the exocrine glandular tissue undergoes atrophy. The muscles responsible for operating a joint which has been immobile for some time undergo marked atrophy and the bones also are affected. Unless such atrophy has become extreme it is reversible and full functional activity may be restored.

3. Interference with the nerve supply. This form of atrophy is seen where there is any destructive lesion of the lower motor neurons or their axons. In this type, **neuropathic atrophy**, there is not only a simple wasting, but also more active shrinkage of the muscle tissue (Fig. 23.68, p.935). For at least a few weeks after nerve section, during which the muscle fibre mass may be reduced by half, anabolic processes take place at a normal rate: catabolism due to increased lysosome numbers and activity is

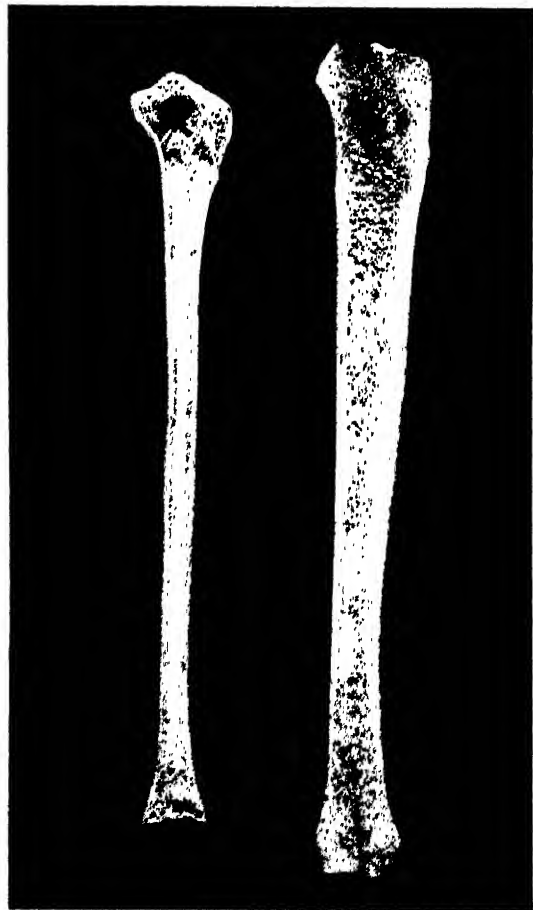


Fig. 2.33 Tibia from a longstanding case of poliomyelitis, showing marked atrophy (*left*). Normal tibia for comparison (*right*). $\times 0.3$.

greatly accelerated. That this form is different in nature from disuse atrophy is shown by the electrical 'reactions of degeneration' which are given by the muscles and indicate that a complete return to normal is no longer possible. Sometimes marked atrophy occurs also in the bones of paralysed limbs; for example, in cases of poliomyelitis the bones of the limb may become thin and light, and this appears to be due simply to inactivity (Fig. 2.33). In some forms of inherited muscular atrophy, however, no nerve lesion is present and the terms *primary myopathy* or *muscular dystrophy* are often applied (p. 933).

4. Deficiency of endocrine glands. Atrophy of thyroid, gonads and adrenal cortex are seen when destruction of the pituitary results in diminished secretion of trophic hormones. In hypothyroidism there occurs marked atrophy of the structures of the skin, hair follicles, sweat glands and sebaceous glands, but structure and function may be restored by oral administration of thyroid hormone.

5. Fever. A good example of atrophy is provided by wasting of the muscles in fevers. No doubt inactivity and loss of appetite play a part, but the wasting is probably due mainly to utilisation of proteins, as is indicated by the increased excretion of nitrogen. This increased protein catabolism is characteristic not only of fever but also follows severe trauma such as fractures or major surgical operations. Other tissues may suffer atrophy in a corresponding way, but in the parenchymatous organs other expressions of cellular injury are more common.

6. Pressure atrophy is also described. The pressure acts mainly by interfering with the blood supply and also with the functions of a tissue. Thus atrophy of the organs may be brought about by the pressure of benign tumours and cysts. When bone is subjected to pressure there is active absorption by osteoclasts.

Examples of atrophy are provided in the later chapters on diseases of the different systems.

Metaplasia

An interesting cellular response to injury is the phenomenon of metaplasia—the transformation of one type of differentiated tissue into another. An example is provided by the surface epithelium of the bronchi which commonly changes from the normal ciliated pseudostratified columnar type to squamous (Fig. 2.34). In this example it appears that chronic injury or irritation, often due to cigarette smoke, results in adaptive changes in the surface epithelium to a type likely to be more resistant to the cause of the irritation. Similarly stratified squamous epithelium may form as a result of chronic irritation in the mucous membrane of the nose, salivary ducts, gallbladder, renal pelvis and urinary bladder. In some cases the injurious stimulus is apparent, e.g. when there is a stone in the renal pelvis or in cases of extroversion of the urinary bladder, while in others the cause is obscure. In vitamin A deficiency, in addition to xerophthalmia, stratified squamous epithelium may replace the transitional and columnar epithelia of nose, bronchi, urinary tract, and the specialised secre-

tory epithelia of the lacrimal and salivary glands. In auto-immune chronic gastritis, in which there is an immunological attack on the mucosa of the fundus of the patient's own stomach, the specialised surface-lining cells and chief and parietal cells of the gastric glands are often replaced by tall columnar cells with striated borders, goblet cells and Paneth cells, i.e. metaplasia to an intestinal type of mucosa.

In the connective tissues, metaplasia occurs between fibrous tissue, myxoid tissue, bone and cartilage. Bone formation occasionally follows the deposition of calcium salts in such tissues as arterial walls (Fig. 2.35), bronchial cartilage and the uveal tract of the eye. In healing fractures cartilaginous metaplasia may occur especially when there is undue mobility. The flattened serosal endothelium of the rabbit pleural cavity becomes cubical, columnar, transitional or even squamous following injection of the dye Sudan III with sodium cholate in olive oil and the lining of adjacent alveoli also becomes cubical or columnar. Similar changes, which



Fig. 2.34 Metaplasia of bronchial epithelium to stratified squamous type is seen on the left side, persistence of columnar ciliated epithelium on the right. $\times 200$.



Fig. 2.35 Metaplastic bone formation in the wall of a largely obliterated artery. $\times 50$.

are rapidly reversible, follow the injection of strontium chloride.

Metaplasia is to be distinguished from a mere loss of the special characters of cells, for example the dedifferentiation which is encountered when there is interference with the function of glands. Developmental epithelial abnormalities, e.g. squamous epithelium within the thyroid, arising from the thyroglossal duct, do not constitute metaplasia, nor does encroachment of one tissue upon another. Thus the fatty marrow of the long bones is replaced in certain types of anaemia by red haemopoietic marrow: in this case the haemopoietic tissue has spread by proliferation of haemopoietic stem cells and not by metaplasia of the adipose tissue cells originally present.

It is believed that all nucleated cells carry a complete list of the genetic information required for bodily development, including all types of cellular differentiation and function, but little is yet known about the factors which determine the differentiation of cells in an orderly

manner to form the various tissues. The way in which the many different stimuli producing metaplasia act within the cell is correspondingly obscure. It seems likely that a change in gene repression and activation takes place in serosal endothelium when it undergoes metaplasia to squamous epithelium. By contrast, in surfaces lined by columnar epithelium, metaplasia may result from gradual atrophy of the columnar cells and proliferation and maturation of the less well differentiated basal or reserve cells to form squamous epithelium. It is noteworthy that many stimuli which bring about metaplasia are also capable of inducing neoplasia, and indeed tumour formation is relatively common in some metaplastic epithelia; conversely metaplasia is frequently encountered in malignant tumours. Indeed metaplasia may represent a cellular change in response to injury intermediate between the kind we have been considering earlier in this chapter and that which underlies the development of tumours.

Further Reading

- Chanarin, I. (1979). *The Megaloblastic Anaemias*, 2nd edn., pp. 1100. Blackwell Scientific Publications, Oxford.
- Cori, G. T. (1952-3). Glycogen Structure and Enzyme Deficiency Glycogen Storage Disease. *Harvey Lecture Vol. 48*, p. 145.
- Dingle, J. T. and Fell, Dame Honor (1973 and 1975). *Lysosomes in Biology and Pathology*, 4 vols. North Holland Publishing Co., Amsterdam.
- Emery, A. E. H. (1979). *Elements of Medical Genetics*, 5th edn., pp. 243. Churchill Livingstone, Edinburgh and London.
- Popper, H. and Schaffner, F. (Eds.) (1976). *Progress in Liver Diseases, Vol. V*, Chapters 4 and 14. Grune and Stratton, New York.
- Stanbury, J. B., Wyngaarden, J. B. and Fredrickson, D. S. (1975). *The Metabolic Basis of Inherited Disease*, 3rd edn., pp. 1778. McGraw Hill, New York and London.

Inflammation

Definition and nature of inflammation

Inflammation may be defined as *a series of changes which take place in living tissue following injury*. While commendably brief, this definition is useless without qualification. We have seen in Chapter 2 how tissue cells may be injured, i.e. rendered abnormal, in many ways, and how the effects may range from pathological storage of metabolites to neoplasia. These are not examples of inflammation, and so it is necessary to qualify both the type of injury and the nature of the changes resulting from it.

The injury which causes inflammation may be brought about by: (1) **physical agents**, such as excessive heating or cooling, mechanical trauma, ultraviolet or ionising radiations; (2) a wide variety of **chemical agents**, both inorganic and organic, and including the **toxins** of various bacteria; (3) the intracellular replication of **viruses**; (4) **hypersensitivity reactions**, i.e. the reaction of antibody or sensitised lymphocytes with antigenic material such as invasive bacteria or inhaled organic dusts; and (5) **necrosis of tissue**, which induces inflammation in the surrounding tissue.

A very important cause of inflammation is **microbial infection**. As indicated above, bacteria produce harmful toxins, viruses injure the host cells which they colonise, and all types of micro-organisms may induce hypersensitivity reactions by the host.

The main features of inflammation. When an appropriate injury, such as excessive heat, is applied to living tissue, an **acute inflammatory reaction** develops. The small vessels in the vicinity of the injury become engorged with blood which at first flows rapidly but gradually slows down. Protein-rich fluid and red cells and subsequently leukocytes escape from the engorged vessels into the tissue spaces. This reaction is

due to changes in the small vessels and, because it includes the escape of blood constituents into the tissues, it is commonly termed **exudative**. When the tissue injury has been slight and brief, the exudative inflammatory reaction is correspondingly mild and soon subsides. However, if the injury persists, the exudative inflammatory reaction can continue for months or even years, as in some persistent bacterial infections, and it is therefore wrong to equate it solely with acute lesions, i.e. those having a short course.

A second type of inflammatory response, sometimes called **productive** or **formative** (to distinguish it from exudative) inflammation, is characterised by proliferation of fibroblasts and production of new fibrous tissue. This occurs particularly in prolonged tissue injury and so is seen especially in **chronic inflammation**. In some instances, both exudative and productive reactions are conspicuous, but in prolonged low-grade injury, fibrous tissue formation is often the more prominent. At first, the young fibrous tissue is highly vascular, soft and gelatinous, and is known as *granulation tissue*. As it ages, it becomes less vascular, progressively more collagenous, and is thus gradually converted to pale, dense *scar tissue*.

The present account follows tradition in using the terms 'acute inflammation' and 'acute inflammatory reaction' for the exudative process, and 'chronic inflammation' for persistent inflammatory lesions, in which fibrous tissue formation is a prominent feature. It must, however, be emphasised that the two types of reaction commonly occur together.

Leukocytes migrate from the blood vessels into the tissues in both acute and chronic inflammation: in the former neutrophil polymorphs usually predominate, while in the latter lymphocytes, plasma cells and monocytes are often more conspicuous.

The inflammatory nature of a lesion is usually indicated by the use of the suffix **-itis**. Thus inflammation of the appendix is **appendicitis**, inflammation of the meninges, **meningitis**, and so on.

Inflammation is usually beneficial. It is essential in combating various infections and in limiting the harmful effects of toxic compounds. Like other beneficial processes, it is not without disadvantages: for example, in acute bacterial infection of the larynx there may be sufficient inflammatory swelling to obstruct the airway and even to cause death from asphyxia, and inflammatory reactions caused by hypersensitivity to harmless substances, as in hay fever,

appear entirely disadvantageous to the host. Fibrous tissue formed in chronic inflammation may help to wall off bacteria or harmful compounds such as silica particles, but it may also cause disability by distorting and compressing important structures. Inflammatory fibrosis of a hollow viscus, such as the intestine, may cause narrowing of the lumen, and fibrosis occurring in any tissue can constrict blood vessels, nerves, ducts, etc.

This chapter deals with the acute inflammatory reaction, the mechanisms involved in its production and its effects, the special features of chronic inflammation, and the types of cell involved in inflammatory reactions.

The Acute Inflammatory Reaction

Acute inflammation has been recognised since the earliest days of medicine. Celsus (30 B.C. to A.D. 38) gave as its cardinal signs **heat, redness, swelling and pain**, to which may be added **limitation of movement**, e.g. of an inflamed limb. The explanation of these features has been provided by microscopic studies, which have revealed that the inflammatory reaction is composed of a number of phenomena, all of which involve the small blood vessels in the inflamed tissue. These phenomena were described by Cohnheim (1889) who observed microscopically the changes in the living transparent tissue of the frog's tongue and foot-web during inflammation caused by mechanical injury or chemical irritation. His superb account is a model of accurate observation and the changes he described have since been confirmed by others in mammalian tissues following thermal or chemical injury. They are as follows.

(a) **Active hyperaemia.** Immediately after thermal or chemical injury there is a transient blanching of the tissue due to arteriolar contraction. This effect, which is not of practical importance, is followed within a few minutes by relaxation of the arterioles in and around the injured tissue, so that the capillary network and post-capillary venules become engorged (Fig. 3.1) with rapidly flowing blood (*active hyperaemia*). This accounts for the **redness (erythema)** of the inflamed tissue, and when the normally cool skin is involved the increased

flow of blood warms it up and so explains the **heat** of inflammation.

(b) **Exudation.** Following the onset of active hyperaemia, protein-rich fluid (the **inflammatory exudate**) escapes from the blood vessels into the surrounding tissues and is largely responsible for the **swelling (inflammatory oedema)**.

The **pain** of acute inflammation is due partly to the rise in tissue pressure resulting from inflammatory oedema; this accounts also for the **relative immobility**, for the oedema increases the rigidity of the tissues and movement further increases the pressure and thus aggravates the pain. Palpation has a similar effect and this explains why inflamed tissues are often exquisitely tender. The original injury (heat, chemicals, etc.) may also be directly painful and, as described later, a number of pain-inducing endogenous compounds (histamine, kinins, prostaglandins, etc.) are released in acute inflammation.

(c) **Slowing of the blood flow.** The microcirculation remains engorged but the blood flow, at first rapid, becomes progressively slower, with even momentary arrests in some of the small vessels.

(d) **Emigration of leukocytes.** Phagocytic leukocytes, at first neutrophil polymorphs and later monocytes, adhere to the endothelium of venules and migrate through the vessel walls into the tissue spaces (Fig. 3.1).

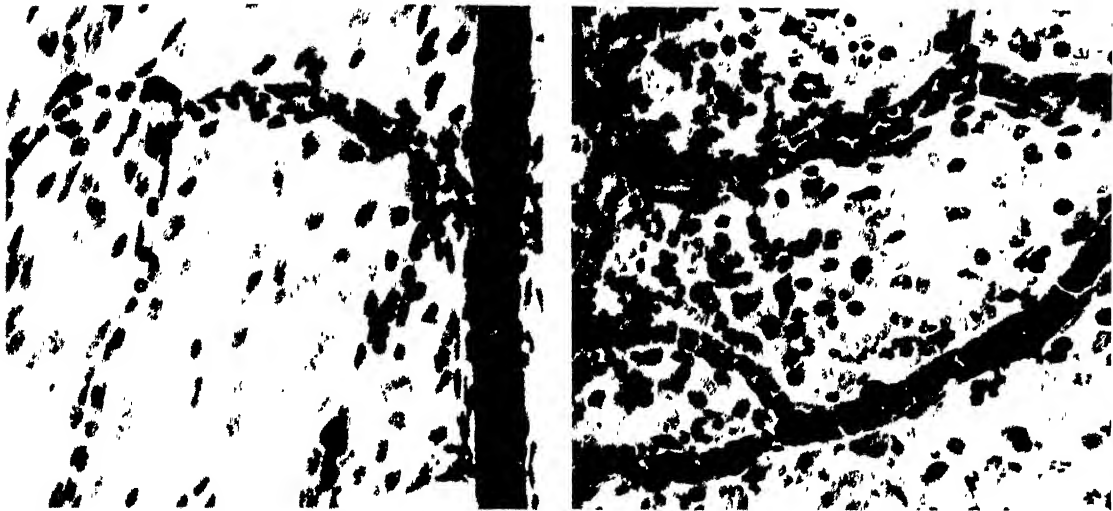


Fig. 3.1 Surface view of stained preparations of guinea-pig omentum showing the normal appearances (*left*) and acute inflammation (*right*). Note engorgement of the small vessels with blood and infiltration of the tissue with neutrophil polymorphs. $\times 300$.

The time of onset, intensity and duration of each of these phenomena vary considerably depending on the type and severity of their causal injury. The reaction to thermal burns and many irritating chemicals is almost immediate; as sunbathers learn to their cost, inflammation of the skin due to excessive ultra-violet irradiation is delayed for several hours, while with ionising radiations, e.g. x-ray, acute inflammation may develop up to a week later.

The features of the inflammatory reaction to microbial infections vary greatly depending on the properties of the microbes and on the host-parasite relationships.

The major phenomena of acute inflammation will now be described in more detail, not only because of their fundamental importance, but also because elucidation of their mechanisms is essential to the advancement of rational therapy.

Active hyperaemia

Normal microcirculatory control

The flow of blood through a tissue is controlled mainly by changes in the tone of the circular smooth muscle of its arterioles. This is regulated in part by the autonomic nervous system, and especially by the sympathetic adrenergic vasoconstrictor nerves, which maintain arteriolar tone and are largely responsible for controlling the blood pressure, the cardiac output, and to some extent the distribution of blood flow among the various tissues. Superimposed on this overall control are local factors determined by the conditions in individual tissues. Thus when an organ or tissue is in a resting state of low metabolic activity, many of its arterioles are contracted and blood flow is dim-

inished. When local metabolism increases, for example in the gastric mucosa after a meal, or in an exercising muscle, accumulated metabolites act directly on the arterioles, causing them to relax and the consequent rise in pressure of blood reaching the capillaries causes them and their draining venules to become engorged with rapidly flowing blood. This engorgement or congestion, due to the rise in pressure of blood entering the capillaries, is termed **active hyperaemia** to distinguish it from the congestion due to a rise in venous pressure (passive hyperaemia) which occurs in other conditions and is not accompanied by an increase in blood flow.

Although the arterioles and venules are greatly dilated in active hyperaemia, the size of capillaries is restricted by their basement membrane: their

diameter is approximately 5–8 μm and in active hyperaemia they become conspicuous not so much by dilatation but because they are filled with whole blood, whereas in the resting tissue state most of them contain slowly flowing plasma with few cells.

In most tissues, the capillaries form an anastomosing network providing routes of various lengths between arterioles and venules. The individual entrances to the network are the *terminal arterioles*, which do not anastomose, and which are the smallest vessels controlled by smooth muscle cells: they function as pre-capillary sphincters and determine the flow of blood through individual capillaries. It is probable that, in resting tissues, the pre-capillary sphincters are so adjusted that blood flows mainly through the shortest capillary routes—the so-called thoroughfare channels—in accordance with the needs of general circulatory control: the pre-capillary sphincters guarding the longer capillary routes are contracted, and most of the capillaries contain only plasma.

The engorgement and rapid flow of active hyperaemia are brought about mainly by relaxation of the arterioles, including most of the pre-capillary sphincters. The capillaries themselves are not contractile and so the amount of blood passing through them depends on the state of the arterioles.

The active hyperaemia of acute inflammation

This is a local phenomenon and is due to pre-dominance of local factors over the general system of arteriolar control. As in physiological activity (see above), the arterioles relax and active hyperaemia results. The mechanism of the arteriolar relaxation is unknown and is difficult to investigate. Either endogenous mediators or neural factors must be involved, for the active hyperaemia extends beyond the immediate site of injury.

In his classical experiments, Lewis (1927) provided evidence for both chemical endogenous mediators and neural factors. He induced mild inflammation in the skin of the human forearm by firm stroking with a blunt point or by mild thermal injury and observed three components in the inflammatory response, namely a *flush* (erythema at the site of injury), a *flare* (erythema of the surrounding skin) and a *weal* (swelling due to inflammatory oedema). Lewis showed that the flush was prolonged by applying a tourniquet to the arm and considered that it was due to capillary dilatation induced by a chemical endogenous mediator, removal of which was dependent on

blood flow. He showed that the erythema caused by a local injection of histamine was prolonged by a tourniquet and called his hypothetical mediator 'H' substance. In subjects with nerve injuries, Lewis showed that the flare could be elicited for up to ten days after severance of the sensory nerve to the part, but not later. Accordingly he postulated that the inflammatory stimulus (firm stroking or thermal injury) triggered off an axon reflex which induced the flare by causing relaxation of the arterioles in the surrounding skin (Fig. 3.2).

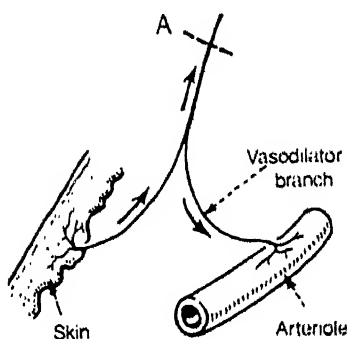


Fig. 3.2 Diagram of peripheral end of sensory nerve fibre with vasodilator axon branch. Stimulation of sensory nerve ending in the skin results in an anti-dromic reflex along the vasodilator branch, with resulting arteriolar dilatation. Section of the nerve fibre at A does not abolish this reflex until the fibre distal to A has degenerated.

It is now known that acute inflammation occurring in long-denervated tissues presents all the usual features and so neural factors clearly play no essential role: it is thus likely that Lewis's postulated axon reflex is not of much practical importance. His work did, however, stimulate considerable interest in the role of endogenous chemical mediators in the acute inflammatory reaction, and an ever-increasing number of endogenous compounds capable of inducing one or more features of acute inflammation has since been detected in inflammatory lesions. The list of possible endogenous chemical mediators of inflammatory hyperaemia includes histamine, 5-hydroxytryptamine, kinins, prostaglandins, products of activation of the complement system, fibrin degradation products and various other polypeptides. These and other vaso-active agents can all induce hyperaemia: what is in doubt is their relative importance in the natural process. To complicate matters further, most of them are capable also of inducing exudation (escape of protein-rich fluid from the small

blood vessels) and yet the hyperaemia and exudation of inflammation do not closely parallel one another either in their timing or in their degree. In inflammation produced in the skin by sunburn, for example, exudation may only occur during a short part of the period of hyperaemia.

Changes in blood flow

The rapid blood flow of active hyperaemia is readily explained by the increased hydrostatic pressure of blood in the microcirculation resulting from arteriolar relaxation.

Slowing of the blood flow. This supervenes gradually until, in some vessels, there may be temporary stasis of blood. In severe inflammation, stasis may be prolonged and clotting may occur. Several factors contribute towards the slowing of the blood flow through inflamed tissues (Fig. 3.3). Firstly, active hyperaemia results in *loss of fluid from the blood* in capillaries and post-capillary venules, and the concentration of cells in the blood is thus increased. Secondly, although the exudate is rich in protein, loss of fluid is so great that there is an *increase in the local concentration of the plasma proteins*. Both of these changes increase the viscosity of the blood locally. Thirdly, the increased protein concentration of the plasma and slowing of blood flow result in *aggregation of the red cells in rouleaux*, with so-called sludging of the blood, and this further increases viscosity. A fourth factor which impairs the blood flow is the *adhesion of leukocytes to the walls of post-capillary venules*. Not only do the leuko-

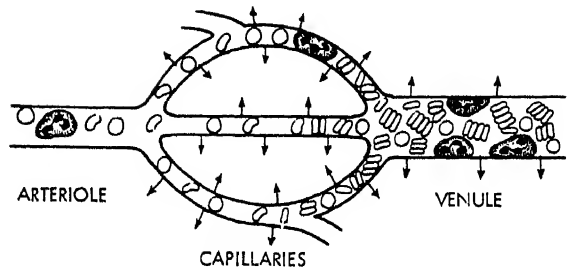


Fig. 3.3 The causes of slowing of the blood flow in acute inflammation. Loss of intravascular fluid (arrows) results in haemoconcentration, with consequent increase in blood viscosity. Increase in plasma protein concentration promotes rouleaux formation by the red cells, further increasing blood viscosity, while pavementing of polymorphs in the venules, together with rouleaux formation, partially blocks the venules. (Modified after Dr Roe Wells; see Zweifach, 1973-4.)

cytes adhere to the endothelium but also to one another, and considerable reduction in the effective lumen of venules may result.

Slowing of the blood flow will tend to impair the supply of oxygen, glucose, etc., to the tissues, and also the removal of metabolites, but these effects are diminished by the increased flow of fluid from the plasma into the tissues and increased lymphatic drainage (see below). It is only when the vascular stagnation is extreme that it is likely to impair tissue nutrition seriously and contribute to the necrosis which is commonly observed in severe inflammatory reactions. Such necrosis is more likely to result directly from the injury which has induced the inflammatory response, e.g. bacterial toxic action, thermal injury, etc.

Exudation of protein-rich fluid

Microscopic examination of inflamed tissues reveals an accumulation of extracellular fluid, i.e. interstitial oedema (Fig. 3.4). This can only have come from the blood plasma, and since it has been shown that the amount of fluid draining away from inflamed tissues by the lymphatics is also greatly increased, there is obviously a considerable rise in the net amount of fluid leaving the blood vessels. As illustrated in Fig. 3.5, the inflammatory exudate is also much richer in plasma proteins than is normal ex-

tracellular fluid (transudate) in the same tissue, indicating increased permeability of the vessels to macromolecules. These two features of exudation—increased loss of intravascular fluid and of plasma proteins—are best considered separately, for the factors involved in the passage of water and small solutes across the walls of microvessels, in both normal and inflamed tissues, differ from those concerned in the escape of molecules as large as the plasma proteins.



Fig. 3.4 Meso-appendix in acute appendicitis, showing inflammatory oedema with early leukocytic emigration. $\times 100$.

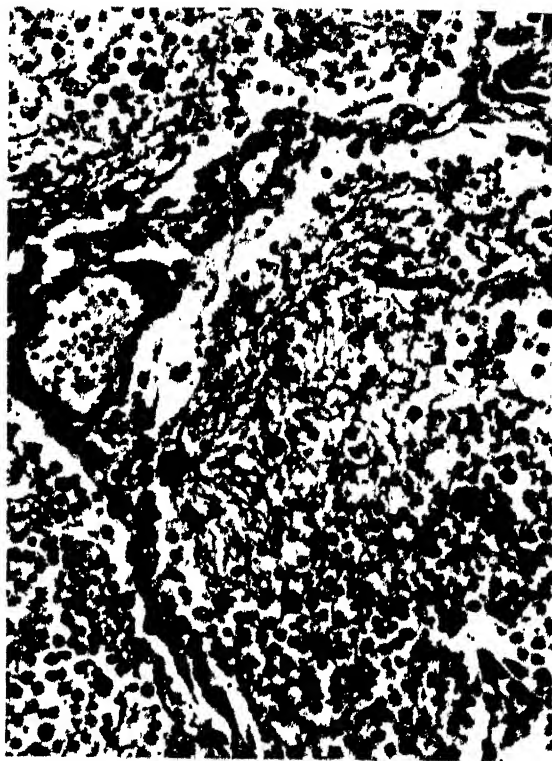


Fig. 3.5 Acute inflammation of the lung in pneumonia. The exudate, which fills the alveoli, is rich in plasma proteins, and this is illustrated by the fine network of fibrin (stained black) which has resulted from the clotting of exuded fibrinogen. (Weigert's fibrin stain.) $\times 150$.

Escape of water and micromolecular solutes

In all tissues the capillaries and post-capillary venules are readily permeable to water and micromolecular solutes. For molecules above a molecular weight of 10 000 daltons, the permeability decreases sharply with molecular size, and molecules greater than 40 000 escape from the plasma in relatively small numbers. This applies particularly to the vessels in skeletal muscles, central nervous system, dermis and other connective tissues. In the liver, intestinal mucosa, exocrine and endocrine glands and the glomeruli, macromolecules escape more readily but still much less so than do micromolecules.

Our understanding of the mass movement of water and small solutes between plasma and extravascular fluid is based upon (a) the observations of Starling and of Landis on micro-filtration, and (b) morphological considerations.

(a) Microfiltration theory

Starling (1896) proposed that the vascular endothelium behaves like a passive microfilter, across which the movement of fluid and electrolytes is determined by physical forces. On this theory the main force driving fluid out of vessels is the height of the hydrostatic pressure within the vessels above that in the extravascular space, and this is opposed by the height of the osmotic pressure of the plasma above that of the extravascular fluid. Thus at the arteriolar ends of capillaries the effective hydrostatic pressure would normally exceed the osmotic pressure and fluid should be forced out. At the venular ends of capillaries the osmotic pressure would exceed the hydrostatic pressure and fluid should be drawn into the capillary (Fig. 3.6). This theory received direct support from the work of an American medical student, Landis (1927), who devised techniques of meas-

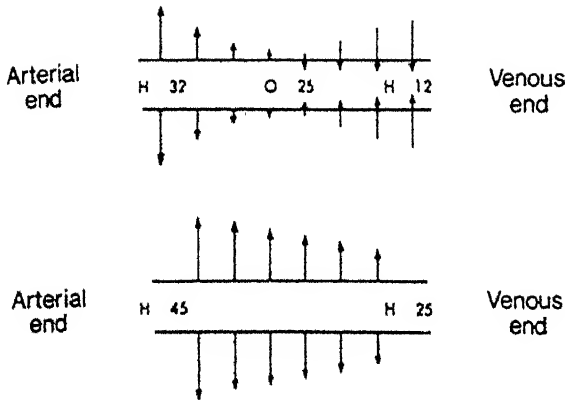


Fig. 3.6 Exchange of fluid across the walls of capillaries and venules. H and O represent the heights of the hydrostatic and osmotic pressure respectively (mmHg) of the plasma above the corresponding pressures of the extravascular space. The arrows indicate the net movement of fluid into and out of the vessels along their length. *Upper figure*, normal tissue: fluid movement across vessel walls approximates to equilibrium. *Lower figure*, acute inflammation: much more fluid leaves the vessels than returns to them.

The values of H and O are approximations. In inflammation, H may be less than indicated because of rise of pressure in the extravascular space, and O will also be reduced by escape of plasma proteins into the inflammatory exudate.

using the hydrostatic pressure in individual small vessels, and of calculating the rate of diffusion of fluid into and out of the vessels. Landis showed that when inflammation was induced in the frog mesentery there was a rise in the hydrostatic pressure within the microcirculation which upset the balance between hydrostatic and osmotic forces, with a resultant net loss of fluid from the vessels. The escape of plasma proteins from inflamed vessels will tend to reduce the osmotic pressure difference between plasma and extravascular fluid and this may be sufficient to accentuate the loss of fluid (Fig. 3.6). These findings have now been confirmed in normal and inflamed mammalian tissues and, while Starling's views have required to be modified in detail to take account of the fluid exchange function of the post-capillary venules, his major suggestion—that *the walls of small vessels behave like a passive filter through which the exchange of fluid is determined by opposing haemodynamic and osmotic forces*—has been widely accepted.

(b) Morphological evidence

Acceptance of Starling's microfiltration theory and subsequent studies have suggested the existence, in the walls of microvessels, of a physiological system of small 'pores' of sufficient size to allow the escape of water and electrolytes, but impermeable to proteins and other macromolecules (Pappenheimer *et al.*, 1951).

There is strong evidence that the basement membrane of small vessels does not form a barrier to water or small molecular solutes and the effective filter thus appears to be in the endothelial layer. Electron microscopy of vascular endothelium suggests two possible routes of fluid transport across the endothelium. Firstly, endothelial cells contain small cytoplasmic vesicles (*micropinocytotic vesicles*), some of which open onto the inner or outer surface of the cell (Fig. 3.7). Simionescu *et al.* (1975) have shown that these vesicles may link up to form channels across the cytoplasmic barrier, through which peptides of below 2000 daltons are capable of passing. Constrictions in these channels may represent the small pores of the endothelial microfilter.

Secondly, fluid may escape *between endothelial cells*. Electron microscopy shows endothelial cell 'junctions' to be potential spaces containing amorphous material (Fig. 3.8). These spaces show constrictions which may act as the endothelial small pores.



Fig. 3.7 Electron micrograph of part of a normal capillary endothelial cell in which micropinocytotic vesicles are seen in relation to both the inner and outer surfaces and also lying free in the cytoplasm. The capillary lumen is at the top of the field. $\times 50\,000$.



Fig. 3.8. Electron micrograph showing the narrow space (arrow), filled with amorphous material, between adjacent vascular endothelial cells (E). L is the lumen of the vessel. $\times 68\,000$.

There appears no doubt that fluid and micro-molecular solutes can cross the endothelium by both the above routes but their relative importance is still uncertain.

Leakage of proteins from microvessels

While relatively permeable to water and small solutes, the walls of normal capillaries and venules exert a sieving effect on larger molecules, providing an increasingly effective barrier to macromolecules proportionate to their size. In connective tissues and voluntary muscle there is relatively little loss of albumin and even less of the larger plasma proteins.* Nevertheless some protein does escape from the microvessels of all tissues, and physiologists have postulated the existence of a second micro-filter system of smaller numbers of larger 'pores' to account for this (Pappenheimer *et al.*, 1951). In an exercised limb, increased pressure and flow of blood through the small vessels of the muscles results in greatly increased flow of the lymph, and yet the total protein content of the lymph, which represents most of the protein escaping from the microvessels, is not in-

creased. Active hyperaemia does not, therefore, account for the increased protein leakage from the microvessels in acute inflammation, which can be explained only by assuming an increase in the number of Pappenheimer's postulated large 'pores'.

The observation of leakage from individual vessels and the detection of Pappenheimer's large pores through which proteins leak out was accomplished by inducing acute inflammation and injecting intravascularly a suspension of carbon particles. The particles escape through the large pores and so label the leaking vessels, while the sites of leakage can be detected by electron microscopy. Using this technique, it was shown by Majno *et al.* (1961) that, when increased permeability was induced by histamine or other potential endogenous mediators (see below), the carbon particles escaped through large gaps which developed between adjacent endothelial cells (Fig. 3.9.). This has now been fully confirmed in acute inflammation induced in various tissues by various injurious agents. The gaps appear to be temporary, for injected colloidal material has also been observed deep to normal (i.e. 'tight') endothelial cell junctions in acutely inflamed tissue. Very occasionally, this has been observed in normal (non-inflamed) tissues, suggesting that transient gaps account also for the normal leakage of small amounts of protein.

Although plasma proteins of various molecular sizes appear in the acute inflammatory exudate, a sieving effect persists, the smaller proteins escaping more readily than the larger ones. Since the observed inter-endothelial cell gaps are much larger than the largest protein molecule, it is apparent that vascular basement membrane also acts as a relatively coarse filter to proteins which have leaked through the endothelium.

It is possible that some protein is normally transferred across the vascular endothelium in the micropinocytotic vesicles (see Fig. 3.7), but these have not been shown to increase in inflammation and are unlikely to be an important factor.

*In small vessels in many other tissues, e.g. glomeruli, glandular tissues, gut mucosa, the endothelial cells show zones of extreme thinning (fenestrae) while the vascular sinusoids of liver, bone marrow, etc., have relatively large endothelial defects. These features probably account for the relatively greater permeability of vessels in these tissues. Studies on acute inflammation have usually been made on the skin and voluntary muscle and are valid only for the 'continuous' (non-constricted) vessels of these tissues.



Fig. 3.9 Two gaps caused by histamine in the endothelium (E) of a venule (rat cremaster muscle, three minutes after a local injection of histamine). Red blood cells are very plastic and can easily 'flow' into endothelial gaps; this one is going to have a problem, because it is slipping out of two different gaps! Note the tight folds in the endothelial nucleus (N) (suggestive of cellular contraction). Because the basement membrane (BM) acts as a filter (beyond the endothelial gaps), many blood-borne particles accumulate in the venular wall: the dark granules are carbon black (India ink which had been injected intravenously); the larger, smooth, round bodies are chylomicrons. (P: pericyte.) $\times 29\,700$. (Dr. Isabelle Joris.)

In summary, exchange of fluid and small solutes across the endothelium of capillaries and venules occurs by passive filtration either through the intercellular material or along channels formed by microvesicles which can bridge the endothelial cytoplasm. The increased leakage of fluid and electrolytes in acute inflammation is explained by the increased hydrostatic pressure of the blood in the small vessels in active hyperaemia; it depends on increased filtration pressure and not on increased permeability. Exudation of plasma proteins from the small vessels in acutely inflamed tissue requires an increase in permeability of the endothelium and this is provided by the reversible opening up of relatively large gaps between endothelial cells: such gaps probably explain also the physiological leakage of small amounts of proteins from normal vessels.

Phases of increased vascular permeability

In experimental studies of increased vascular permeability, acute inflammation has usually been induced by readily controlled injury, e.g. mild heat or injection of irritating chemicals. Increased permeability may be observed by use of colloidal carbon as mentioned above, but is more readily measured by administering an intravascular injection of either a dye such as Evans blue which binds to plasma proteins, or of plasma albumin labelled with a radioactive isotope. Blue discolouration of the tissue or accumulation of the isotope then indicates escape of protein from the blood vessels. By such techniques, it was shown by Seivitt (1958) that moderate thermal injury of the skin of a guinea-pig results in two phases of increased vascular permeability (Fig. 3.10). The first phase is *immediate and transient*, subsiding within about thirty minutes. The second phase is *delayed and prolonged*, starting after one to two hours, reaching a peak in about four hours and subsiding slowly. Milder thermal injury results only in the immediate transient phase, while severe injury induces *immediate persistent* increase in permeability continuing for twenty-four hours or more. These patterns, with some variations, have been demonstrated using thermal, chemical and bacterial toxic injury (Burke and Miles, 1958) to induce inflammation in various tissues and animal species.

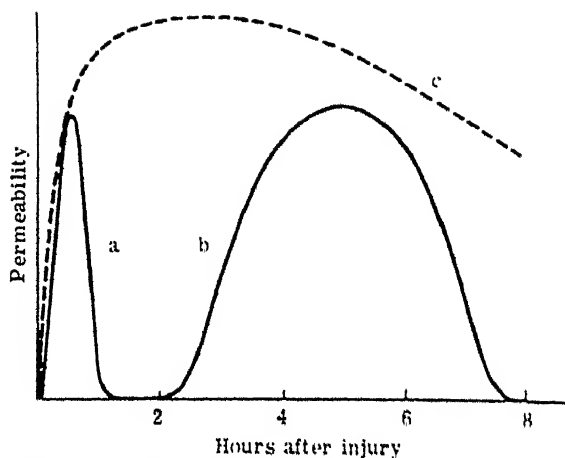


Fig. 3.10 Phases of increased vascular permeability following injury to the skin by application of heat. Following mild injury, there is immediate but transient increase in permeability (a); with moderate injury, this is followed by delayed prolonged increase in permeability (b), while severe injury causes immediate persistent increase (c). (Modified after the late Professor D. L. Wilhelm: see Zweifach, 1973 4.)

Mechanisms of increased vascular permeability

The phases of increased vascular permeability described above have been further analysed by electron microscopy using colloidal carbon as a 'vascular marker' (p. 50) to indicate the sites of escape of proteins. During the **immediate transient phase**, leakage occurs through the endothelium of venules only (Fig. 3.11). Since all potential endogenous mediators so far investigated (e.g. by use of carbon marking) have been found to increase the permeability of venules, but not of capillaries, this finding suggests that the immediate transient phase is endogenously mediated. Moreover, when the zone of injury is sharply defined, as is the case when a hot metal tube is applied to the skin, the increased venular permeability of this phase is seen to extend to the surrounding uninjured tissue, (Fig. 3.12), which suggests diffusion of endogenous mediators. A third feature is that the endothelial cells of the leaking venules present changes like those described by Majno (see Ryan and Majno, 1977a) after injection of histamine: the cells appear plumper and their nuclei become more rounded and show crinkling of the nuclear membrane (Fig. 3.9). These features suggest that the endothelial cells contract, becoming shorter and fatter, and pull apart from one another, thus accounting for the observed gaps between cells. It is noteworthy that histamine and most of the other

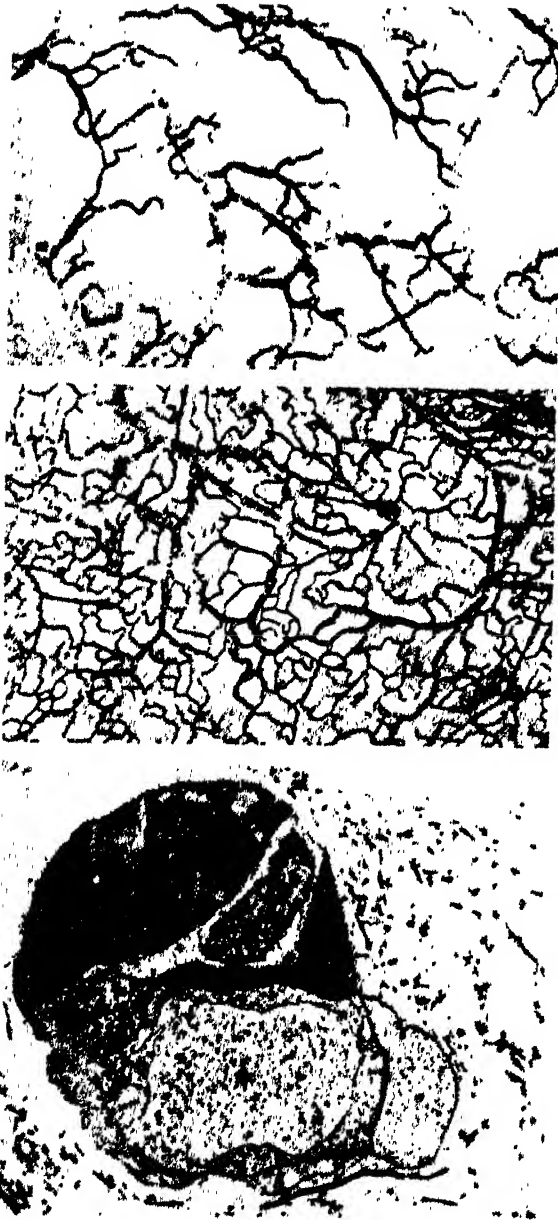


Fig. 3.11 Carbon labelling of vessels to demonstrate increased vascular permeability following thermal injury. *Upper*, the immediate transient phase, showing labelling of venules only. *Middle*, the delayed persistent phase, showing labelling of venules and capillaries. *Lower*, a dermal capillary showing endothelial injury: the lumen is marked by an asterisk. Note leakage of carbon. (By permission of Dr Guido Majno and the Upjohn Company.)

potential mediators of increased permeability are known to cause contraction of smooth muscle, e.g. in isolated intestine or uterus. A few years ago, the suggestion that endothelial cells can contract like smooth muscle cells would have

seemed heretical, but endothelial cells (and also polymorphs, monocytes and platelets) are now known to contain contractile proteins resembling actin and myosin and cytoplasmic micro-fibrils resembling those of smooth muscle cells. Accordingly, it seems likely that *endogenous mediators cause increased venular permeability by stimulating endothelial cells to contract*.

By contrast, in the **delayed persistent phase** of increased permeability induced by application of heat, *there is leakage of macromolecules from both venules and capillaries (Fig. 3.11) confined strictly to the zone of tissue injury (Fig. 3.12)*. This observation was made by Hurley (1972) who concluded that this phase was not due to endogenous mediators but to *the delayed effect of direct injury on the endothelium*, a view supported by electron-microscopic evidence of endothelial injury (Fig. 3.11c). This conclusion is of considerable importance, for it follows that attempts to inhibit excessive prolonged leakage of plasma proteins, e.g. in extensive burns, should aim at diminishing the effects of endothelial injury rather than at the suppression of endogenous mediators. Hurley made the interesting observation that promethazine, which reduces the delayed injurious effect of carbon tetrachloride on the liver cells, also diminished the delayed prolonged phase of increased permeability following thermal injury. The **immediate prolonged increase in permeability** following more severe injury presents features

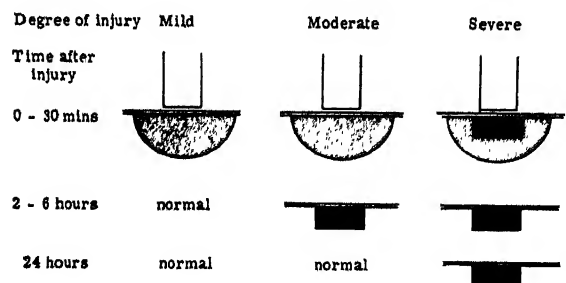


Fig. 3.12 The phases of increased vascular permeability following various degrees of heat injury to the skin by a hot tube. Two effects are observed. Firstly, an immediate transient increase in venular permeability (lightly shaded), due to release of histamine, etc.: this probably occurs with all degrees of injury short of necrosis and extends beyond the zone of injury. Secondly, more persistent injury (heavily shaded) results from direct heat injury to the endothelium of capillaries and venules: this is delayed in moderate injury and immediate in severe injury, and is confined to the zone of injury.

similar to those of the delayed prolonged phase, and is due to a greater degree of direct endothelial injury. Thrombosis may occur in the badly damaged vessels, but otherwise exudation continues until the damaged endothelial cells recover or are replaced. In injury severe enough to cause tissue necrosis, exudation does not, of course, occur in the dead tissue, but immediate persistent exudation occurs in the severely injured adjacent tissue.

Observations similar to those described above have been reported in experimental inflammation induced in various animal species by diverse injurious agents, e.g. x-irradiation, ultraviolet light, certain bacterial toxins and crushing injury. There are, however, exceptions: for example, the delayed persistent increase in permeability which occurs when carrageenan is applied to the cremaster muscle of the rat is confined to venules and appears to be endogenously mediated. Also, different tissues behave differently: for example, various types of injury to the skin of the rat result in increased permeability apparently due to direct injury, but the same agents applied to the rat diaphragm induce prolonged increase in permeability apparently due to endogenous mediators.

In summary: exudation of protein-rich fluid in acute inflammation has been shown experimentally to occur in two phases—one immediate and transient, and the other delayed and prolonged. The immediate phase is due to release of histamine and other endogenous mediators which appear to cause contraction of venular endothelial cells, with consequent development of temporary gaps between them, through which protein-rich fluid escapes. Endogenous mediators may also play a role in the delayed phase, but in some instances there is strong evidence that the direct effect of the causal injury upon the capillary and venular endothelium is responsible for the intercellular gaps observed in both types of vessel in this phase.

Endogenous mediators of increased vascular permeability

Although the work of Hurley and others has demonstrated that increased vascular permeability can and does result from the direct effects of injury on the vascular endothelium, it is apparent from the above account that endogenous

mediators also contribute. A large and increasing number of endogenous compounds which can increase vascular permeability have indeed been demonstrated in effective concentrations in inflamed tissues and in inflammatory exudates. To prove that a particular compound is responsible for increased permeability it would be necessary also to demonstrate that specific suppression of its production or activity results in a reduction in increased permeability. This is difficult to achieve, for potential endogenous mediators are numerous and complex, and any procedure or antagonist which suppresses one is liable to interfere with others. Moreover, mediators differ in their effects on different animal species. For example, 5-hydroxytryptamine is much more effective in increasing venular permeability in rats and mice than in other mammalian species. However, the immediate transient phase is now widely regarded as being mediated largely by histamine for it is partly inhibited by low doses of relatively specific histamine antagonists, e.g. mepyramine maleate, and by prior histamine depletion, e.g. by Polymixin B. Apart from this, none of the potential mediators has yet been shown conclusively to be of practical importance in increasing vascular permeability. Accordingly, the following notes on their production and properties are intentionally brief.

Mediators derived from the plasma

These include various products of activation and interaction of four major 'cascade' systems—the clotting, fibrinolytic, kinin and complement systems. Each system has a number of components which include pro-enzymes, conversion of which to the active enzymes can trigger off the activation of subsequent components in the system, giving a chain or cascade reaction. Each system is complicated by the presence in the plasma of inhibitors and accelerators, and by positive and negative feedback mechanisms. Moreover some of the activation products of the individual systems can interact with the other systems. Fig. 3.13 is a gross over-simplification of these complexities. The clotting and fibrinolytic systems are described in more detail on pp. 234–5 in relation to thrombus formation and lysis, and the complement system on p. 142 in relation to its activation by antigen-antibody complexes.

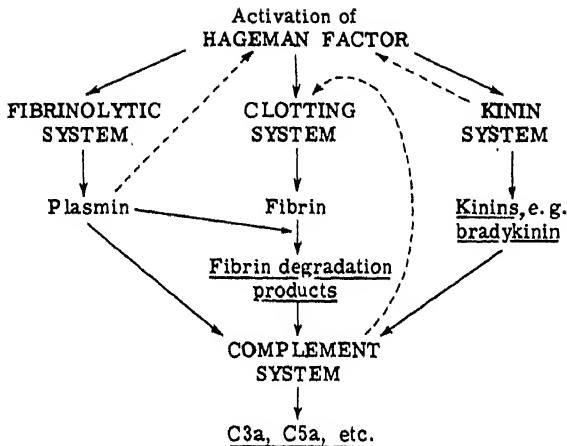


Fig. 3.13 A simplified version of the complex interactions between the fibrinolytic, clotting, kinin and complement systems. The first three of these systems are triggered off by activated Hageman factor and products of their activation activate more Hageman factor. The complement system can be activated by products of the fibrinolytic and kinin systems. Products underlined cause active hyperaemia, increase the permeability of venular endothelium, and some are also chemotactic for leukocytes. Controlling and inhibitory factors are not shown in the diagram.

As shown in Fig. 3.13, **Hageman factor** in the plasma (traditionally factor XII of the clotting system) plays a key role, for once activated it can initiate activation of the clotting, fibrinolytic and kinin systems, each of which can, in turn, activate more Hageman factor. *In vitro*, Hageman factor is activated by contact with negatively-charged surfaces, e.g. glass or kaolin. *In vivo*, it is activated by contact with various extracellular tissue elements, including basement membrane, by bacterial endotoxin (p. 178) and by various proteolytic enzymes of the kinin, clotting and fibrinolytic systems. It is thus apparent that, once inflammation has developed, Hageman factor leaking through endothelial gaps can become activated by contact with basement membrane and activation of three of the major systems can ensue. Products of the activity and interactions of these three systems induce inter-endothelial cell gaps, apparently by causing endothelial cell contraction: they include fibrinopeptides, fibrin degradation products and kinins. In addition (Fig. 3.13), activation products of the fibrinolytic and kinin systems can activate the complement system which also generates permeability-increasing factors and which, in turn, can activate the clotting system. As mentioned earlier,

these permeability-increasing factors are also capable of promoting active hyperaemia (p. 46).

In established inflammatory lesions, large numbers of enzymes are released by tissue cells and leukocytes. These include proteases, some of which are capable of activating components of the four major plasma systems shown in Fig. 3.13, and also of breaking down tissue and plasma proteins into peptides, some of which can themselves act as permeability-increasing factors.

The **kinin system** was revealed largely by the work of Miles and Wilhelm (1955). They detected a permeability-increasing factor (termed PF/dil) which was generated spontaneously from diluted plasma in a glass container. The system has been elucidated mainly by the work of Cochrane and Wuepper and their colleagues (Cochrane *et al.*, 1974). The main components of the system are shown in Fig. 3.14. Prekallikrein activator is a product of activated Hageman factor. The other components of the system are all present in the plasma, but enzymes with kallikrein activity are present also in most tissues and in urine and glandular secretions. Bradykinin is a nonapeptide derived from breakdown of kininogen, a plasma glycoprotein, by the proteolytic action of kallikrein. On injection, bradykinin causes pain, active hyperaemia and increased venular permeability. Several closely related peptides have similar properties: they are rapidly destroyed by kininases in the plasma and tissues, which also contain kallikrein antagonists. The latter are commercially available for therapeutic use.

The **complement system** is described more fully in relation to hypersensitivity on p. 142-4. It consists of a series of components termed C1, C2, etc., in the plasma. Activation of the system generates agents which increase venular permeability. These are termed anaphylatoxins and include C3a and C5a which are cleavage products of C3 and C5

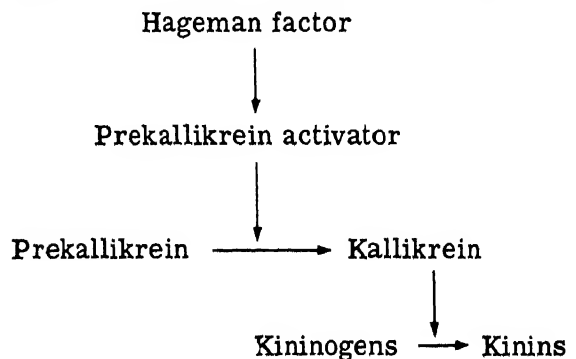


Fig. 3.14 The major components participating in the cascade reaction of the kinin system.

respectively. They act mainly by liberating histamine from mast cells, but C5a is also strongly chemotactic for leukocytes (p. 60). Plasma contains a potent inactivator of these products (which may also inactivate kinins).

In the complex environment of the inflammatory reaction there are a number of ways in which C3a and C5a may be generated: the main ones are as follows.

- (a) In *bacterial infections*, the host's plasma may contain antibodies to antigenic components of the invading bacteria (perhaps developed during a previous infection): the union of such antibodies with the corresponding antigens can activate the complement system.

The endotoxins of Gram-negative bacteria can activate the complement system at the C3 stage. Also, some bacteria secrete proteases which are capable of cleaving C3 and C5.

- (b) *Tissue injury* itself can also activate complement, for the reaction of auto-antibodies (often present at low levels in the plasma) with the various corresponding tissue antigenic components released by tissue injury provides antigen-antibody complexes capable of activating complement.

When injured, heart muscle, and probably other tissues, also release enzymes capable of cleaving C3 and C5.

- (c) *Neutrophil polymorphs and macrophages* which have migrated into the tissues and are phagocytosing tissue fragments or bacteria, etc., secrete lysosomal enzymes, some of which are capable of cleaving C3 and C5.
- (d) As indicated above, some products of activation of the other plasma cascade systems (clotting, fibrinolytic and kinin systems) are capable of activating the complement system.

Although the above observations do not establish that complement plays a part in non-immunological inflammatory reactions, this possibility is suggested also by the demonstration that animals depleted of plasma complement by various methods show impaired inflammatory exudation following physical or chemical tissue injury. Since such animals react normally to injected histamine, kinins, etc., complement appears to be of importance either as a source of mediators or by stimulating the production of mediators from other sources (Willoughby, Coote and Turk, 1969; Lewis and Turk, 1975).

Mediators released by cells

These are also numerous, but only a few have been characterised chemically. They include the following.

Histamine is stored in inactive form in the granules of *mast cells*, which are present adjacent to blood

vessels in most tissues, and also in basophil leukocytes and platelets. Active histamine is released from these cells by many substances and stimuli, including those which induce acute inflammation, e.g. heat, irradiation, irritating chemicals, toxins, venoms, by anaphylatoxins (p. 55), and by a lysosomal protein secreted by polymorphs in inflammatory exudates. Release of histamine from platelets is also stimulated by factors which cause platelet aggregation (p. 232). Its release in some types of immunological reaction is described in Chapter 6.

On local injection, histamine causes itching and pain, active hyperaemia and increased venular permeability for approximately fifteen minutes, after which the vessels are said to become refractory for several hours to a further injection of histamine. Compounds which deplete the mast-cell store of histamine, and others which inhibit the venular response to histamine by competitive binding to endothelial-cell receptors, do not have important anti-inflammatory effects in low dosage, although they partly inhibit the immediate transient phase of increased vascular permeability.

5-hydroxytryptamine (serotonin) is present in most tissues. Rich sources include the cells of the chromaffin system of the gastro-intestinal tract, the spleen and nervous tissue, mast cells and platelets. On injection, 5-hydroxytryptamine, like histamine, causes a brief increase in venular permeability: it may participate in inflammatory reactions in rats and mice, which are particularly sensitive to it, but in other species, including man, it is unlikely to make an important contribution to inflammation.

Prostaglandins are a group of long-chain hydroxy-fatty acids which are produced in most tissues by the action of an oxidase (PG-synthetase) on polyunsaturated fatty acids such as arachidonic acid. They are rapidly catabolised and are not stored within the body. While they differ greatly in their properties, prostaglandins E1 and E2 have been isolated from inflammatory exudates in man and animals and shown to be capable of causing active hyperaemia, increased vascular permeability and possibly chemotaxis of polymorphs. They are also potent pyrogens when injected into the third ventricle and although small doses intradermally do not cause pain, they lower the pain threshold of nerve endings to histamine, 5-hydroxytryptamine and kinins. Control of their synthesis and release is obscure, but they have been detected in inflammatory exudates and are secreted by phagocytically active polymorphs. Firm evidence that prostaglandins play an important part in inflammation is scanty, but it is of interest that aspirin and related drugs, which have anti-inflammatory, anti-pyretic and analgesic properties, have been shown to be capable, in low concentrations, of inhibiting the production of prostaglandins, both *in vivo* and *in vitro*, by antagonising prostaglandin synthetase activity.

Neutrophil polymorphs and monocytes. As stated earlier, these cells migrate from the venules into the tissues in acute inflammation, and become more actively motile and phagocytic. During this activity, and also when these migrated cells are injured by bacterial toxins, etc., they release various lysosomal enzymes and other proteins, many of which participate in the activation of the plasma cascade systems (p. 55) or break down various plasma proteins into fragments which increase venular permeability. In addition, polymorphs secrete a factor which stimulates the release of histamine, etc. from mast cells and other factors which act directly on venules, increasing their permeability.

Conclusions. *The early (immediate transient) phase of increased venular permeability observed*

in mild inflammation is due partly to histamine release. There are many potential mediators of the later phase of increased venular permeability, but none has been implicated with certainty. It seems reasonable to conclude that the important process of exudation has been safeguarded by the development, during evolutionary selection, of multiple mediators of increased venular permeability, no single one of which is indispensable.

Preoccupation with endogenous mediators should not be allowed to obscure the evidence provided by Hurley and others that, in some experimentally-induced inflammatory reactions, prolonged increase in permeability affects capillaries and venules and is due to direct endothelial injury (p. 53).

Emigration of leukocytes

The escape of cells from the blood vessels is a prominent feature of inflammation. Escape of erythrocytes is purely passive: they are forced out of capillaries and venules, through gaps between endothelial cells, by the hydrostatic pressure of the blood. Their escape in very large numbers is an indication of severe endothelial injury. By contrast, escape of **neutrophil polymorphs and monocytes** is an active process of great importance, and of particular significance in the defence against bacteria. It is independent of the endothelial gaps responsible for increased vascular permeability and involves two stages: firstly, the leukocyte becomes arrested on the surface of the vascular endothelium, and secondly it passes through the vessel wall.

In acute inflammatory lesions, neutrophil polymorphs migrate earlier and in much greater numbers than monocytes.

Margination of polymorphs

Arrest of neutrophil polymorphs on the vascular endothelium is often conspicuous in acute inflammation and is known as **pavementing or margination** of leukocytes. It is seen solely in venules and occurs with the slowing of the blood flow in the dilated vessels. In the earlier stage of rapid flow, blood in the arterioles and venules shows **axial streaming**, the cells being

mainly in the central or axial columns of blood, separated from the vessel wall by a clear layer of plasma containing only occasional cells. This streaming is dependent on the rapid flow of blood and later, as the rate of flow decreases, axial streaming disappears. In particular, the leukocytes in the venules pass into the peripheral stream, where they can make contact with the endothelium. Neutrophil leukocytes making such contact tend to become arrested momentarily and then become detached and move on, or roll slowly along the endothelial surface. Eventually more and more of them become arrested for longer periods on the endothelium and they may form an almost continuous layer or may even become heaped up on one another (Fig. 3.15). The nature of adhesion between the leukocytic and endothelial cell surfaces is unknown: changes in the cell surfaces have not been detected by electron microscopy.

In recent studies, vascular endothelial injury in small vessels has been caused by a fine laser beam (5–15 μm diameter) during perfusion with saline coloured with a dye, and adhesion of leukocytes to the injured endothelium has been observed following restoration of blood flow. Since blood cells and plasma were excluded from the vessel during injury, pavementing can clearly result from endothelial injury alone, and does not require injury to the leukocytes.



Fig. 3.15 Section of venule in acute inflammation, showing pavementing of polymorphonuclear leukocytes. $\times 1000$.

Emigration of polymorphs

The 'pavemented' polymorph pushes out cytoplasmic pseudopodia and when one of these encounters the junction between two endothelial cells, it extends between them, disrupting the junction (Fig. 3.16), and the rest of the cell squeezes through: the intercellular junction reforms rapidly without significant leakage of plasma. The emigrating polymorph also passes through the basement membrane, which is repaired almost immediately. The mechanisms of disruption and repair of the endothelial cell junction and basement membrane are not known. Polymorphs take only a few minutes to pass these barriers; they then wander through the tissues and play a role in digestion and phagocytosis of fibrin, degenerate tissue and cell fragments and, most important in infections, in the destruction and removal of micro-organisms. The phagocytic function of leukocytes is considered on pp. 63 and 180.

Intensity of polymorph emigration in acute inflammation depends upon the nature and severity of the tissue injury. It is usually only moderate in physical injury unless infection supervenes. Inflammatory chemicals, including bacterial products, vary greatly in the degree of leukocytic emigration they induce. In the mild inflammatory reaction which occurs around tissue dying from acute ischaemia, i.e. an infarct (p. 246), the degree of polymorph emigration also varies greatly. In myocardial infarction, for example, there may be virtually no polymorph infiltration of the dead muscle, or large numbers may be present, particularly near the margin.

The outstanding examples of intense emigration of polymorphs are provided by bacterial infections: bacteria which, like *Strep. pyogenes*, *Staph. aureus* and *Strep. pneumoniae* are particularly active in this respect, are accordingly termed **pyogenic** (pus inducing) bacteria. Other bacteria, such as, *Salmonella typhi* (the cause of typhoid fever) and *Clostridium welchii* (a cause of gas gangrene), induce far less polymorph emigration, even though they cause severe inflammation. These special features are considered in more detail below and in Chapter 8, but it is worth noting here that differences between inflammatory reactions are not simply in degrees of severity: the nature of the injurious agent determines to some extent the relative degrees of the various features (exudation, emigration of leukocytes, etc.) of the reaction.

Chemotaxis

The migration of polymorphs through the walls of venules and their subsequent movement in the tissues has been widely assumed to be mediated by chemotaxis, a process in which cells move towards higher concentrations of certain substances termed **chemotactic agents** or **chemotaxins**. Such directed movement is not readily demonstrated *in vivo*, largely because it is difficult to establish and maintain gradients of concentration of test substances in living tissues. Nevertheless, time-lapse cinephotomicrography of inflamed tissues within rabbit ear chambers* has revealed that the movements

*The rabbit ear chamber consists of two thin, flat transparent plastic discs with a narrow space between and open around the edge. For use, it is sutured in a round hole punched in the pinna, and a layer of vascular connective tissue grows in to occupy the space. Inflammation can be induced in the connective tissue by various means and the changes examined microscopically *in vivo* (Fig. 4.5, p. 80).



Fig. 3.16 A neutrophil polyporph caught in the act of emigrating out of a venule (rat omentum; experimental inflammation caused by sterile necrotic kidney tissue). Note the many fibrils in the endothelial cell at top; above it, part of a platelet (P1)—which must have slipped out earlier (not necessarily through the gap now being used by the neutrophil). (E: Endothelium; P: Pericyte; P1: Platelet.) $\times 24\,470$. (Dr. Guido Majno.)

of polymorphs in pursuit of bacteria appear as purposeful as a dog following a scent.

Two methods have been used extensively to detect chemotactic agents. In the method of Boyden (1962) a suspension of leukocytes is separated by a millipore membrane from the test solution. If the latter contains chemotactic agents, the leukocytes migrate through the pores of the membrane (Fig. 3.17) and measurement of their rate of advance can provide a reliable assay of chemotactic activity. Precau-

tions must be taken, however, to exclude the effects of **chemokinetic agents**, which enhance random motility of leukocytes without influencing their directional motility. In the other method (Harris 1953), the suspension of leukocytes is incubated in a slide-coverslip preparation in the presence of a source of the agent being tested for chemotaxis; the movement of individual leukocytes is observed microscopically, usually by time-lapse cinephotomicrography.



Fig. 3.17 Electron-micrograph of a neutrophil polymorph migrating through a millipore membrane in response to a chemotaxin. Most of the organelles have passed into the cytoplasm which, together with two lobes of the nucleus, has moved downwards through a pore, the site of which is indicated by the heavy line. $\times 14\,300$. (By courtesy of Dr. P. C. Wilkinson and Churchill-Livingstone.)

Chemotactic agents for neutrophil polymorphs

By use of the above methods, it has been shown that there are many substances which are chemotactic for neutrophil polymorphs *in vitro*. They include products of (a) the complement, clotting, fibrinolytic and kinin systems, (b) injured tissues, (c) micro-organisms, (d) polymorphs during phagocytic activity, and (e) partial digestion of proteins in the inflammatory exudate.

(a) **The plasma cascade systems.** Activation of the clotting, kinin and fibrinolytic systems, already discussed in relation to the mediation of increased vascular permeability (p. 54) results in chemotactic products. These include *fibrinopeptides*, *fibrin degradation products* and *kallikrein*.

Activation of the complement system also

produces chemotaxins, the strongest of which is C5a (a cleavage product of C5): a complex of activated C5, 6 and 7 may also be chemotactic. Complement seems to be of particular importance in chemotaxis and accordingly its activation in inflammatory lesions, already considered on p. 55, is further commented on below.

(b) **Tissue injury.** Injured tissue cells release compounds which are directly chemotactic for polymorphs, including *prostaglandin E1* (which is also a potential mediator of increased vascular permeability), and enzymes which activate the complement system. The importance of complement as a source of chemotactic agents when tissue injury occurs is suggested by experiments in which necrosis of part of the myocardium was induced in rats by ligation of a coronary artery: emigration of polymorphs from the blood vessels at the margin of the infarct was found to be largely suppressed by prior depletion of the plasma C3 (Hill and Ward, 1971).

(c) **Micro-organisms.** Some bacteria secrete lipid or protein compounds which are directly chemotactic for polymorphs, for example, *Staph. aureus* and *Esch. coli*.

As described on p. 56, the reaction of microbial antigens with host antibodies can result in activation of complement, while bacterial endotoxins can directly activate complement at the C3 stage. Some bacteria also secrete proteolytic enzymes capable of cleaving C3 or C5.

Although viruses are not known to produce chemotactic agents, cells containing replicating virus have been shown to do so.

(d) **Neutrophil polymorphs** which have migrated into inflamed tissue secrete an agent which is chemotactic for other neutrophil polymorphs. Once they have become actively phagocytic, polymorphs secrete lysosomal enzymes, some of which are capable of cleaving C3 or C5. During phagocytic activity, polymorphs also secrete a substance which immobilises other polymorphs in the vicinity. These cells thus appear to have the means to ensure their continued replacement in an inflammatory lesion so long as tissue debris, bacteria, etc. are available for phagocytosis.

(e) **Partly denatured proteins.** The inflammatory exudate contains various proteolytic enzymes derived from injured tissue cells, from migrated polymorphs and from activation of the plasma cascade systems. Proteins in the exudate or released by cell injury are thus exposed to mild proteolysis, the effect of which in some instances (e.g. plasma albumin, immunoglobulin G and haemoglobin) is to render them chemotactic.

The above account, by no means comprehensive, may serve to indicate the complexity of chemotactic agents. These have been

detected by *in-vitro* experiments, and the role of chemotaxis *in vivo* is not established. It is, however, very likely that it is important in the accumulation of leukocytes in inflammatory lesions, for a correlation has been shown between the chemotactic properties of various compounds and their *in-vivo* capacity to induce migration of polymorphs.

Conclusions. *In their number and complexity, chemotactic agents in acute inflammatory lesions resemble potential mediators of increased vascular permeability. Some agents have both properties, e.g. prostaglandin E1, complement products and fibrinopeptides. However, increased vascular permeability and emigration of polymorphs occur independently, often from different vessels and at different times during the inflammatory response. This is not surprising in view of the important role of direct endothelial injury in increased permeability. (p. 53.)*

The mechanism of chemotaxis

Chemotactic agents for polymorphs are many and diverse, and so it is unlikely that recognition of each depends on the existence of specific receptor sites on the cell surface. Wilkinson (1974) has suggested that hydrophobic chemical groups are important in conferring chemotactic properties, and this would explain why mild denaturation or proteolytic digestion of some proteins renders them chemotactic, for such treatment exposes hydrophobic groups. This proposal is supported also by the demonstration that coupling of non-polar (hydrophobic) groups to proteins renders them chemotactic, while polar groups are without this effect.

Polymorphs which have responded to a chemotactic agent become refractory to this and other chemotactic agents, although they can still phagocytose and kill bacteria. In explanation, it has been proposed that chemotactic agents operate by activating the enzyme serine esterase, which is present in inactive form on the cell surface. Once activated, the enzyme decays rapidly and is apparently not regenerated by the polymorph. In support of this possibility, agents which inhibit serine esterase suppress chemotactic responses. In preliminary experiments, Wilkinson *et al.* (1977) have made use of chemotactic agents labelled with a fluorescent dye. They have shown that such agents bind diffusely to the surface of responding cells, but after an hour or so the agent aggregates at one part of the cell surface and is then ingested by the cell. These changes are accompanied by loss of chemotactic responsiveness and immobilisation of the cell: they are consistent

with removal from the cell surface of serine esterase or some other agent necessary for chemotaxis.

Movement of leukocytes is probably effected by the contraction of myofibrils formed by polymerisation of actin and myosin and attached to the inner side of the plasma membrane. Wilkinson (1974) has proposed that direction of movement is controlled by the cell's microtubule system and this is supported by the demonstration that colchicine, which inhibits the polymerisation of tubulin to form microtubules, renders the cell incapable of chemotactic responses without interfering with their random motility.

In most acute inflammatory lesions, **eosinophil polymorphs** emigrate from the vessels in relatively small numbers. They are reported to respond chemotactically to some bacterial products and to cleavage products of C5. Their behaviour in inflammation caused by allergy is described on p. 148.

Very few **basophil leukocytes** are observed in most acute inflammatory reactions.

Emigration of monocytes

Neutrophil polymorphs emigrate earlier and more rapidly than monocytes, so that in short-lived acute inflammation the peak of polymorph emigration has passed before monocytes emigrate in significant numbers. In more prolonged inflammation due to pyogenic bacterial infection, emigration of polymorphs continues until most of the bacteria have been destroyed, and only then do monocytes emigrate in large numbers. It is thus apparent that different factors control emigration of polymorphs and monocytes. Chemotactic responses of monocytes have not yet been investigated extensively, but it is apparent that both monocytes and macrophages are attracted by some of the agents which are chemotactic for polymorphs, for example some bacterial products, cleavage products of C5 and kallikrein.

In inflammation due to infection with some bacteria, e.g. *Myco. tuberculosis* and *S. typhi*, emigration of polymorphs is transient or absent, and most of the emigrating cells are monocytes and lymphocytes. The role of cell-mediated immunity in such responses is discussed in Chapter 7, but monocyte emigration also predominates in the experimentally-induced reaction to relatively inert foreign material, such as carrageenan and synthetic polymers, which are unlikely to invoke an immunological

reaction. It is thus apparent that there are chemotactic agents which predominantly attract monocytes. One such is a lysosomal protein secreted by polymorphs during phagocytic activity, while a product of *Corynebacterium parvum* appears to be specifically chemotactic for monocytes: an intradermal injection of this bacterium induces emigration mainly of monocytes.

Lymphocytes and chemotaxis. There is recent evidence that lymphoid cells can respond chemotactically to some agents, while they are capable also of influencing migration and motility of polymorphs, monocytes and other lymphocytes (Chapter 6).

The lymphatics in acute inflammation

The smallest lymphatics are blind-ending tubes

with a very thin endothelium and a fine, incomplete, i.e. discontinuous, basement membrane. Normally they are partly collapsed, but fine fibrils attach the outer surface of the endothelium to the collagen in the surrounding tissue, and swelling of the tissue by inflammatory exudate tenses these fibrils and *distends* the lymphatics. The endothelial cells overlap one another and their junctions are very easily separated: they appear to act as valves, allowing fluid to pass in but not out.

These features allow greatly increased lymph drainage from inflamed tissue. Exuded proteins are removed by the lymphatics, and red cells and leukocytes also pass into the lymphatics of inflamed tissue.

The filter function of the lymph nodes in inflammation is described in Chapter 18.

Effects of acute inflammation

Acute inflammation is classed as a pathological process although there is no doubt that its effects are, in general, beneficial. It helps to eliminate invasive micro-organisms, to limit the injurious effects of irritating chemicals and bacterial toxins, and participates in the removal of necrotic cells and tissue debris.

Like most beneficial biological processes, acute inflammation is not without its disadvantages: in some instances it appears to confer no obvious benefit, and in others it seems positively harmful.

Beneficial effects

These are conferred partly by the flow of exudate through the inflamed tissues and partly by the phagocytic and microbicidal effects of emigrated leukocytes.

The inflammatory exudate

The fluid exudate is protective in the following ways.

1. Dilution of toxins. When inflammation is caused by toxic chemicals, including bacterial toxins, the exudate diminishes local tissue injury by diluting the toxins and carrying them away by the lymphatics.

2. Protective antibodies. The proteins in the exudate include antibodies which have developed as a result of infection or immunisation and which are present in the individual's plasma. In acute inflammation due to infection, the exudate may thus contain antibodies which react with, and promote destruction of, the micro-organisms, or which neutralise their toxins. Antibodies promote killing of micro-organisms by rendering them susceptible to lysis by complement and destruction by phagocytes. This is described more fully in Chapter 7.

3. Fibrin formation. Fibrinogen in the exudate is converted to solid fibrin by the action of tissue thromboplastin. A network of deposited fibrin is commonly seen in inflamed tissues, and may form a mechanical barrier to the movement and spread of bacteria. It may also aid in their phagocytosis by leukocytes.

4. Promotion of immunity. Micro-organisms and toxins in the inflammatory lesions are carried by the exudate, either free or in phagocytes, to the local lymph nodes where they may stimulate an immune response. This provides antibodies and cellular mechanisms of defence which appear within a few days and may be maintained for years.

5. Cell nutrition. The flow of inflammatory

exudate brings with it glucose, oxygen, etc., and thus helps to supply the greatly increased numbers of cells: it also carries away their metabolic products.

Phagocytosis

The neutrophil polymorphs in inflammatory lesions are actively phagocytic. The emigrated monocytes are not at first so active, but they rapidly change into the larger, more active macrophages. The process of phagocytosis is similar for both polymorphs and macrophages, and resembles closely the engulfment of food particles by amoebae. First, the surface of the phagocyte attaches to the particle, e.g. bacterium, to be ingested. The cytoplasm then flows around the particle and envelops it in a **phagocytic vacuole**. Finally the plasma membrane enclosing the vacuole breaks away from the cell surface, and the membrane-lined vacuole lies free in the cytoplasm. The subsequent fate of the particle depends on its nature and on the host's response. Adjacent lysosomes fuse with the membrane of the phagocytic vacuole, and pour their contents into it, the vacuole now being termed a **phagolysosome** or **phagosome**. The particle is thus exposed to the lysosomal acid hydrolases, and these include such a wide range of enzymes that most biological material, including red cells, fibrin, collagen and ground substance, dead cells and cell components, are digested. By engulfing and digesting the debris of the inflammatory reaction, the phagocytes act as scavengers (Fig. 3.18). During phagocytic activity, polymorphs and macrophages also release lysosomal enzymes into the surrounding fluid where they contribute to the digestion and so removal of inflammatory debris: the digestion products include peptides, nucleotides, etc. which, by increasing vascular permeability and attracting leukocytes by chemotaxis, may enhance the inflammatory reaction.

Polymorphs and macrophages play a vital protective role in microbial infections. In most bacterial infections, the bacteria are eliminated rapidly by phagocytosis and other protective mechanisms. However, there are exceptions, and some micro-organisms live and even multiply in phagocytes. The factors concerned in these host/parasite relationships are considered in Chapter 7.

Neutrophil polymorphs are highly specialised cells; they are actively motile, rich in lysosomal enzymes, and respond to relatively early chemotactic stimuli in the inflammatory reaction. They have a rich store of glycogen, and enzyme systems which provide the energy required for motility and phagocytosis by glycolysis. The last property allows polymorphs to function in the low oxygen tension present in highly cellular inflammatory exudates.

Because of these properties, the polymorph is admirably suited to its role in early defence against acute bacterial infections: it arrives early on the scene and is aggressive in engulfing and killing bacteria. It is, however, an end-stage cell and is unable to re-synthesise the surface plasma membrane and granules (lysosomes) used up in these activities. In consequence it soon loses its granules, becomes ineffective and dies. The supply of polymorphs is, however, practically unlimited.

Monocytes are less actively motile and phagocytic than polymorphs. They provide a reserve of cells which, on emigration in an inflammatory lesion, change into macrophages: this involves increases in lysosomal enzymes, metabolic activity, motility, and phagocytic and microbicidal capacity. Like polymorphs, they have enzyme systems which supply the energy for this increased activity by anaerobic glycolysis, but they differ in having little stored glycogen and must therefore make use of glycogen released by polymorphs or glucose in the exudate as a source of energy. Macrophages can ingest and destroy inflammatory debris (dead cells, fibrin, tissue fragments, etc.—Fig. 3.18) and can envelop and sequester indigestible material, e.g. foreign bodies and certain micro-organisms, for long periods: they can synthesise plasma membrane, lysosomal enzymes and lysosomes and are capable of division and of long survival after phagocytic activity. These properties suit them particularly to sustained function in prolonged inflammatory reactions.

Harmful effects

Swelling of acutely inflamed tissues may have serious mechanical effects. For example, in acute laryngitis the lumen of the larynx may be so reduced as to interfere with breathing.

Acute inflammation of tissues which are

confined within a restricted space, and so cannot expand, results in a rise of tissue pressure which may impair function directly or may interfere with blood flow and so cause ischaemic injury. Examples include inflammation of the brain (encephalitis) and meningitis, both of which cause increased intracranial pressure sometimes leading to coma and death. Similarly acute bacterial infection of the bone marrow (osteomyelitis) raises the pressure in the medullary cavity and extensive ischaemic necrosis may occur. A third, painful example is acute inflammation of the testis, usually caused by mumps virus; the tough tunica albuginea prevents much expansion, and ischaemia results,

sometimes with permanent residual injury. Fortunately, both testes are seldom severely affected.

Some examples of inflammation are inappropriate and harmful. For example, most people encounter grass pollen in the air without ill effect, but others become sensitised to pollens and react by the acute conjunctivitis and rhinitis of hay fever. There is also a rare condition, angio-neurotic oedema, in which acute inflammatory lesions develop spontaneously in various tissues, including the gastro-intestinal tract. It is due to deficiency of a plasma factor which controls activation of complement, and the inflammatory lesions are apparently mediated by complement activation products.

Further stages of acute inflammation

The three common results of the acute inflammatory reaction are:

- (a) Resolution, i.e. subsidence of the inflammatory changes and return of the tissue to normal.
- (b) Progression to suppuration.
- (c) Progression to a chronic phase with fibrosis.

Resolution

Termination of the injury which has caused acute inflammation is followed by reversal of the inflammatory changes, and provided that there has not been wholesale destruction, the tissue usually returns to normal. Cell and tissue debris are digested by enzymes in the exudate or by phagocytes (Fig. 3.18); pavementing and emigration of leukocytes cease; the vessel walls regain their normal permeability, and blood flow returns to normal. Most of the emigrated polymorphs probably die, while macrophages (emigrated monocytes) may pass to the draining lymph nodes. Inflammatory exudate drains away in the lymphatics, and normality is restored.

A striking example of resolution is presented by lobar pneumonia, an acute infection of the lung usually due to *Streptococcus pneumoniae*, in which typically the alveoli throughout a whole lobe become filled with a protein-rich exudate containing a fine network of fibrin and large numbers of neutrophil polymorphs (Fig.



Fig. 3.18 A late stage of pyogenic bacterial infection, showing phagocytosis of polymorphs, red cells and debris by macrophages. $\times 510$.

16.24, p. 468). Following destruction of the bacteria, usually after several days, polymorphs and macrophages complete the digestion of fibrin, dead cells and debris, the fluid exudate is removed partly by reabsorption and partly by coughing, and in most cases the lobe returns to normal.

Suppuration

Pyogenic bacteria (p. 58) cause acute inflammation in which emigration of polymorphs is intense, and in which local toxic injury is often severe enough to cause tissue necrosis at the centre of the lesion. The dead tissue is digested by the polymorph enzymes, leaving a space in the tissue filled with inflammatory exudate rich in polymorphs (Fig. 3.19) and containing also

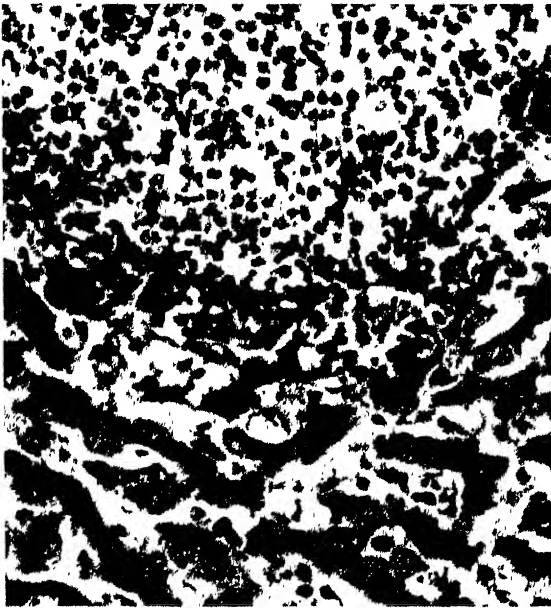


Fig. 3.19 Margin of an abscess cavity in the liver. In the upper part of the field the liver tissue has been destroyed and digested, leaving a space filled with purulent exudate. $\times 250$.

bacteria, fragments of necrotic tissue, cell debris and fibrin. Such a cavity is termed an **abscess**, and the contained fluid, which may be creamy from its cell content and sticky from its high nucleic acid content (from dead polymorphs) is called **pus** or **purulent exudate**. Return to normal is no longer possible since tissue has been destroyed, and the abscess becomes enclosed in a wall of granulation tissue (the pyogenic membrane) which eventually matures to scar tissue. Pus can also form in a natural body cavity, such as the pleura or peritoneum, without tissue destruction, as a result of pyogenic bacterial infection.

A more detailed account of suppuration and its effects is given in Chapter 8.

Fibrosis in acute inflammation

Although acute inflammatory lesions frequently subside without leaving any significant residual changes, this is by no means always so. Formation of granulation tissue with consequent fibrosis or scarring is a common result. It complicates acute inflammation when there is necrosis of tissue or excessive deposition of fibrin and when acute inflammation persists and becomes chronic.

Tissue necrosis. Inflammation has been defined as the reaction to injury of *living* tissue, but many injuries, e.g. burns or bacterial infections, bring about necrosis of tissue. Obviously, inflammatory changes cannot occur in necrotic tissue, but inevitably the adjacent, surviving tissue is injured less severely and inflammation occurs in it. Accordingly, necrotic tissue is commonly present in the centre of acute inflammatory lesions. The occurrence of such wholesale necrosis of tissue, as distinct from necrosis of single cells, precludes the possibility of return to normal. If the dead tissue is superficial, as in a burn, it



Fig. 3.20 A dense layer of fibrin on the pleural surface, showing organisation, i.e. replacement by vascular granulation tissue, extending from the underlying pleura. $\times 250$.

usually becomes detached, leaving a gap in the surviving tissues. If deeper, it may be gradually replaced by granulation tissue (a process termed **organisation**), or it may be digested, as in suppurating infection, leaving an abscess cavity. In all three instances, granulation tissue grows from the adjacent living tissue, and matures to fibrous scar tissue.

It is worth noting that when tissue is excised, leaving a gap, or dies from ischaemia (lack of blood supply), its place is usually taken by fibrous tissue in the process of healing. The fibrosis which follows necrosis of tissue in acute inflammatory lesions is thus an example of healing.

Organisation of fibrin. The inflammatory exudate contains plasma proteins, including

fibrinogen, and frequently this is converted to insoluble fibrin which is deposited in the inflamed tissue. Fine strands of fibrin (Fig. 3.5, p. 48) are readily digested by proteolytic enzymes in the exudate or removed by phagocytosis (p. 63). Larger deposits of fibrin, however, are not readily removed in this way, but, like dead tissue, are more gradually replaced by granulation tissue (Fig. 3.20) by the process of organisation, with consequent scarring. This is commonly seen in acute inflammation of a serous membrane, such as the pleura or pericardium (Fig. 3.21), when a thick layer of fibrin is deposited on the surface, and its subsequent organisation results in fibrous thickening.

Chronic inflammation. Progression of acute to chronic inflammation is described below.

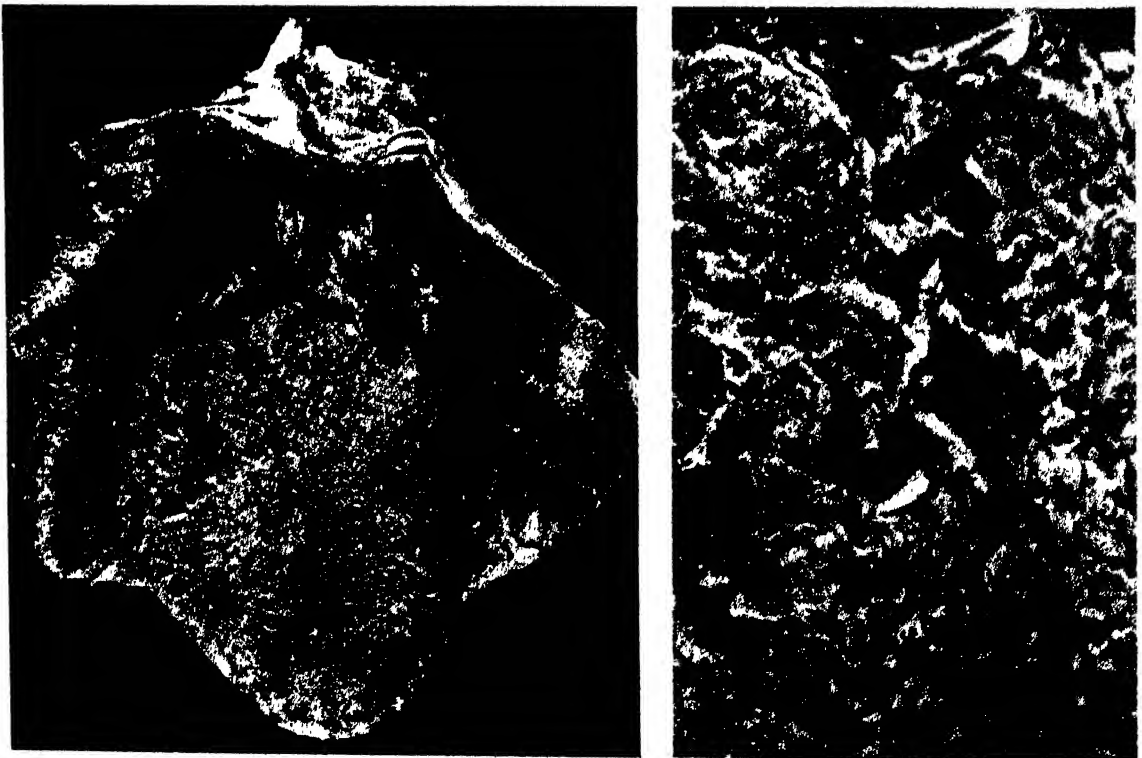


Fig. 3.21 Acute pericarditis, showing a thick, irregular deposit of fibrin on the pericardial surfaces. *Left*, $\times 0.7$; *right*, $\times 3$.

Chronic Inflammation

In contrast to tissue injury of short duration, which induces a brief or acute inflammatory reaction, prolonged tissue injury causes persistent, or chronic inflammation. There is, how-

ever, no generally accepted time limit beyond which an inflammatory lesion is regarded as chronic. To some extent, it depends on the nature of the disease process concerned: for

example a whitlow lasting for several weeks might well be regarded as chronic, as compared with the usual short course, while tuberculous lesions showing extensive spread within weeks are regarded as acute in contrast to those which smoulder on for months or years.

An important feature of chronic inflammation is the production of vascular granulation tissue, which matures into fibrous tissue. Such **proliferative changes** are in contrast to the exudative changes of the acute inflammatory reaction, but when acute inflammation fails to resolve and becomes chronic, the two processes are commonly associated. Even in an early stage of acute inflammation, some proliferation of fibroblasts occurs and, in general, the more chronic the inflammatory reaction, the more pronounced are the proliferative changes, and the greater the degree of ultimate fibrous scarring.

Causes and types of chronic inflammation

Inflammation is a response to tissue injury and soon subsides when the causal injury ceases. Acute inflammation is elicited by relatively intense injury which is usually brief, but in some instances it may persist, although in less intense form, and there is progression from acute to chronic inflammation. There are also many agents, both microbial and inanimate, which cause tissue injury which is prolonged, but of low grade from the start.

Chronic inflammation may thus have an acute onset, or may develop more insidiously.

Chronic inflammation with an acute onset

The outstanding example of this is *bacterial infection*. Most acute bacterial infections are rapidly eliminated, and the tissue returns to normal or, if there has been tissue destruction or excessive fibrin deposition, replacement fibrosis and scarring ensue. But in some instances the infection is only partly subdued in the acute stage, the bacteria survive in relatively small numbers and so the inflammation progresses to a chronic stage. This is exemplified by bacillary dysentery (infection of the colon), in which

there is an acute exudative reaction in the mucosa of the colon. This may resolve or may progress to chronic inflammation in which the exudative changes are accompanied by patchy tissue destruction with ulceration of the mucosa, formation of granulation tissue and scarring. Another important example is provided by pyogenic infection of a bone which, unless treated early and effectively, is likely to cause extensive necrosis of the bone. Bacteria persist within the dead tissue, where they are protected from the host's defence mechanisms, and the infection may continue for years.

It is particularly in this form of chronic inflammation that the bacteria survive in foci and continue to promote an exudative reaction with emigration of polymorphs. From time to time the infection may flare up, with abscess formation. Such infected foci become enclosed in granulation tissue which may form large masses, and which, as it ages, becomes converted into dense fibrous tissue. The granulation tissue is infiltrated with polymorphs, as in the acute stage, but also with macrophages, and with lymphocytes and plasma cells which reflect the host's specific immune response to the infection.

Chronic inflammation of insidious onset

Agents which cause low-grade but persistent tissue injury, and thus promote this form of 'primary' chronic inflammation, fall into the following main classes.

(a) **Particulate material.** This is phagocytosed by macrophages. If it is bland, e.g. suture material (Fig. 3.25), or carbon dust deposited in the lungs from polluted air, there is little or no fibrous reaction and the change can scarcely be regarded as inflammatory. More irritating particles, such as silica, may also be inhaled or may enter the tissues in dirty wounds or in the form of talc formerly used to lubricate surgical gloves.* This also stimulates phagocytosis by macrophages, but the silica dissolves very slowly within the phagolysosomes, yielding silicic acid which injures the macrophages in such a way that they secrete lysosomal enzymes and eventually die. The supply of macrophages is, however, maintained by emigration of monocytes and, on

*Starch powder, currently in general use to lubricate surgical gloves, has also been reported to cause peritoneal inflammation in some cases following abdominal operation.

release from dead cells, the particles are ingested by fresh ones. These events are accompanied by formation of abundant dense fibrous tissue. The stimulus to fibrogenesis is not understood, but Allison has suggested that it is induced as a consequence of leakage of lysosomal enzymes (see Allison, 1978). He has demonstrated that various materials which induce such leakage when phagocytosed by macrophages in cell culture induce fibrosis *in vivo*.

The inflammatory response to fibrous silicates (asbestos) is similar to that of silica. Some other substances, e.g. particles of beryllium compounds, also promote chronic inflammatory lesions in which it seems likely that a state of hypersensitivity is involved (see below.)

(b) Microbial infections. Various bacteria and fungi cause chronic infections of insidious onset. Some of these organisms are remarkably non-toxic but can live and multiply within macrophages without destroying them. Unless the host develops a state of hypersensitivity to the organisms, the infected tissues become heavily infiltrated with macrophages containing huge numbers of the organisms. There is little or no fibrosis but such lesions are classed as chronic inflammation, partly because they are infections and partly because the host develops an antibody response which is reflected by the presence of lymphocytes and plasma cells, formerly regarded as 'chronic inflammatory cells', in and around the lesions. Examples include lepromatous leprosy (Fig. 8.20, p. 215) and leishmaniasis (p. 564).

Other bacteria, exemplified by the tubercle bacillus, are equally non-toxic but the features of the lesions are modified by the development of a state of **delayed hypersensitivity** by the host. This not only promotes killing of the bacteria by macrophages, but results also in tissue injury with necrosis, granulation tissue formation and scarring. The hypersensitivity reaction is a result of the host's immune response and is characterised by infiltration of the lesions with lymphocytes and increased numbers of macrophages which adopt a characteristic appearance and arrangement (Fig. 8.13, p. 209). Similar changes are seen in the tuberculoid form of leprosy and in the reaction to the eggs of schistosomal worms (Fig. 20.46, p. 698).

In many chronic infections, e.g. syphilis,

there is tissue destruction, formation of granulation tissue and aggregation of macrophages, lymphocytes and plasma cells, but it is not yet known how much the destructive changes are due to toxic injury by the causal organism and how much they are the result of the host's hypersensitivity.

(c) Hypersensitivity reactions. The important part played by hypersensitivity reactions in some chronic infections is noted above. Hypersensitivity to normal tissue constituents also occurs and is responsible for chronic autoimmune thyroiditis, gastritis, etc. Such lesions produce important effects by destruction of the parenchyma of the affected organs.

A form of chronic inflammation of the skin, termed *contact dermatitis*, is also a manifestation of a hypersensitivity reaction: it occurs when various substances are absorbed into the skin, where they react with, and modify, host proteins; in consequence, the host develops immunity and a state of delayed hypersensitivity to the affected skin, and this results in an inflammatory rash which may be chronic if exposure to the chemical responsible for it is prolonged. A good example is the nickel used in thimbles and in the fasteners of women's underwear: slight solution of the nickel occurs on sweating and nickel compounds are absorbed and react with epidermal proteins.

(d) Unknown agents. In some chronic inflammatory diseases the causal agents remain unknown. Important examples are *sarcoidosis* (p. 216) which produces lesions somewhat similar to tuberculosis, and *ulcerative colitis* (p. 622) which may necessitate removal of the colon.

The appearances of chronic inflammatory lesions

From the examples given above, it will be apparent that chronic inflammation presents considerable histological variety. The features include foci of acute exudative inflammation, sometimes with active suppuration, infiltration with polymorphs, macrophages, lymphocytes and plasma cells, necrosis, and formation of granulation tissue and dense fibrous tissue. (Figs. 3.22—3.25, 3.28) Variations result from the relative prominence of each of these features, and on the site and distribution of the causal agent. Moreover, macrophages can present various appearances, not only dependent



Fig 3.22 Chronic inflammation: showing the confusing variety of cells. In this instance, they include plasma cells, lymphocytes and occasional polymorphs, while the larger nuclei probably belong to macrophages and fibroblasts. $\times 820$.

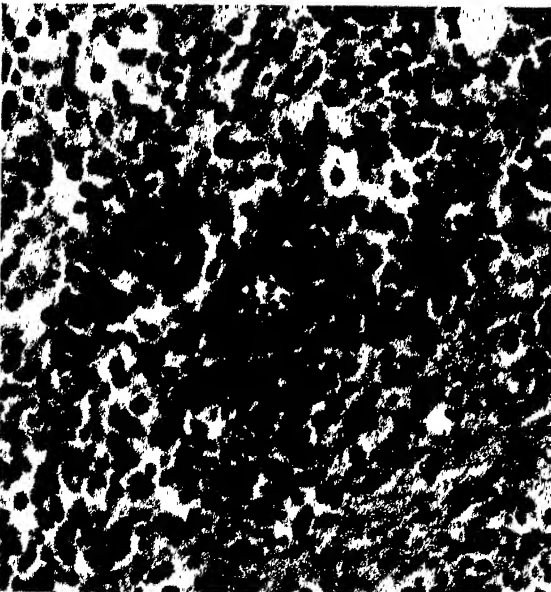


Fig. 3.23 Chronic inflammation: showing newly formed vascular fibrous tissue which is heavily infiltrated with macrophages (top left) lymphocytes and plasma cells. $\times 300$.



Fig. 3.24 Chronic inflammation. This field shows vascular fibrous tissue infiltrated mainly with lymphocytes. $\times 560$.

on the numbers which accumulate, but also on their arrangement and adoption of striking morphological features (p. 73).

The macroscopic appearances are equally variable: abundant granulation or fibrous tissue can lead to tumour-like swellings or more diffuse enlargement; in other instances, loss of parenchyma and scarring are predominant, the tissue becoming firm and shrunken. If due to pyogenic infection, there may be sinus formation with leakage of pus from underlying abscesses, and even without suppuration, necrosis may occur and extend to the skin or a mucous membrane, with consequent ulceration.

In some examples, the gross and/or microscopic features are sufficiently characteristic to suggest a specific diagnosis, but there are dangerous pitfalls: for example, several agents can give rise to changes readily mistaken for tuberculosis. Accordingly, *a firm diagnosis usually depends on the recognition of a specific causal agent, or on other procedures such as serological tests for a particular infection*. Sometimes specific bacteria or fungi can be detected in the lesions by their morphology and staining

properties, while anisotropic foreign material can be detected and sometimes identified by microscopic examination in polarised light (Fig. 3.25).



Fig. 3.25 Foreign-body giant cells in a chronic inflammatory reaction to suture material, viewed through partly crossed polarising films to show up the birefringent foreign material. $\times 250$.

Granuloma and granulomatous inflammation

These terms are widely used, but with three different meanings,

Traditionally they are used to describe a chronic inflammatory lesion in the form of a mass and thus grossly resembling a tumour: hence the suffix *-oma* which is usually reserved for true tumours. Such inflammatory masses are usually due to infection (**infective granulomas**): they are usually composed largely of granulation and fibrous tissue, but may contain foci of suppuration, or may consist mainly of aggregated macrophages as in lepromatous leprosy.

Recently there has been an increasing tendency to limit the use of 'granuloma' to lesions

composed largely of macrophages, or even to restrict it further to mean the collection of altered macrophages (epithelioid and giant cells) found in the tubercles of tuberculosis (Fig. 8.13, p. 209) and similar, non-tuberculous lesions. Since tubercles are often barely visible to the naked eye and do not contain granulation tissue, this is an important change of meaning which does, however, seem likely to be generally accepted. Meanwhile, it is advisable to indicate this usage by such terms as **macrophage granuloma**, **epithelioid cell** or **tuberculoid granuloma**.

Finally, a number of heterogeneous entities are called granulomas, e.g. *malignant granuloma of nose* or *midline granuloma*, (a lesion with histological features of chronic inflammation which resembles a tumour in its infiltration and destruction of tissue), *Wegener's granuloma* (due to a polyarteritis) and *eosinophil granuloma of bone* (a mass consisting mainly of eosinophil leukocytes). These terms are distinctive enough to avoid confusion.

Effects of chronic inflammation

When due to an infection, chronic inflammation is a protective process. Granulation tissue forms a barrier to bacteria and their toxins and provides numerous small vessels from which exudation and emigration of leukocytes can continue. In many instances, as explained above, chronic inflammation is due partly to a delayed hypersensitivity reaction to micro-organisms. This is a defensive process which nevertheless causes tissue injury, and it is often difficult to know whether tissue injury is due to the invading bacteria or to the host's reaction to them. When chronic inflammation results from hypersensitivity to otherwise harmless chemicals, e.g. in contact dermatitis, it seems to subserve no useful function, and the same applies to the chronic inflammation of autoimmune thyroiditis, etc.

In some instances of chronic inflammation with features of hypersensitivity reactions, the causal agent is unknown, and although the inflammation causes tissue injury it may conceivably help to suppress a causal micro-organism as yet unidentified. Sarcoidosis (p. 216), Crohn's disease (p. 620) and rheumatoid arthritis (p. 917) fall into this category.

The fibrous tissue which is formed in chronic inflammation may induce serious effects by constricting orifices and tubes—for example, the mitral valve in chronic rheumatic fever (p. 417) or the small intestine in Crohn's disease (p. 620). Chronic inflammation of internal organs is usually accompanied by loss of parenchymal cells, and this, together with irregular fibrosis, results in shrinkage, irregular scarring and distortion. Commonly the surface becomes uneven, with a fine or coarse granularity: this is particularly well seen in cirrhosis of the liver (Fig. 20.31, p. 685), where the irregularity is accentuated by proliferation and enlargement of surviving liver cells.

In some instances, the fibrous tissue produced in chronic inflammation may have a useful function: for example, walling off chronic infections, or strengthening the aorta weakened by loss of muscle and elastic tissue in various forms of arteritis.

Other causes of fibrosis

While fibrosis is an important feature of chronic inflammation, it may result from other causes. As stated above, fibrosis is the usual method of repair when tissue has been lost, and occurs in the removal of deposits of fibrin by organisation. Unless dissolved by fibrinolytic enzymes, thrombus in blood vessels is also replaced by fibrous tissue. These processes are dealt with in the next chapter.

When the blood supply to a part is gradually diminished by arterial disease, atrophy of the specialised cells may be accompanied by overgrowth of the supporting tissue. Similarly, death of tissue resulting from sudden occlusion of an artery, e.g. by thrombosis, is followed by replacement of the dead tissue by fibrous tissue. Patches of fibrosis of this nature are commonly seen in the myocardium. These effects of deficient blood supply (*ischæmia*) are described in Chapter 9.

Types of Cell in Inflammatory Lesions

Polymorphonuclear leukocytes

Neutrophil polymorphs. The origin and morphology of these cells are described on p. 184 *et seq.* Their migratory activity has already been considered and their roles in the defence against micro-organisms and in Arthus type hypersensitivity reactions are dealt with in later chapters.

Eosinophil polymorphs. The accumulation of these cells in inflammatory lesions is closely associated with hypersensitivity reactions, in which they may play a modulating role (p. 148). They are observed particularly in the lesions of bronchial asthma, in the tissues around metazoan parasites, in certain skin diseases, and in various lesions of the gastro-intestinal tract. Intense local accumulation of eosinophils is commonly associated with eosinophil leukocytosis in the blood and there is recent evidence to suggest that this is mediated by an

immune response. The thymic-dependent lymphocytes which respond to antigenic stimulation, e.g. by a parasitic worm, in some way stimulate the proliferation of eosinophil precursors in the bone marrow.

Lymphoid cells

Lymphocytes accumulate in chronic inflammatory lesions and their presence in large numbers is suggestive of either a delayed hypersensitivity reaction or possibly of antibody-dependent lymphocyte cytotoxicity. These phenomena are described in Chapters 5 and 6.

The presence of **plasma cells** in inflamed tissues, as elsewhere, is indicative of antibody production: they do not usually appear until about a week after onset of inflammation and are present in greatest numbers in persistent lesions caused by bacteria. Their origin and function are described in Chapter 5.

Macrophages: the mononuclear phagocyte system

The terms **macrophage** and **mononuclear phagocyte** were applied by Metchnikoff, in 1905, to large phagocytic cells, which he distinguished from the smaller phagocytic neutrophil polymorph. In 1924, Aschoff described investigations on tissue cells based on vital staining, i.e. the ingestion of droplets of fluid (pinocytosis) containing non-toxic dyes bound to protein, and concentration of the dye in the cell cytoplasm. While many cells did this, Aschoff noted particularly intense staining of cells in the lining of vascular and lymphoid sinusoids, reticular cells of the spleen and lymph nodes, and scattered cells lying in connective tissues. He grouped these cells together under the term *reticulo-endothelial system*. From this grouping, the macrophages have emerged as cells which, although widely dispersed through the body, share a common origin and have certain well-defined functions. Accordingly the term 'mononuclear phagocyte system' is being used increasingly for macrophages and their precursor cells. The term 'reticulo-endothelial system' is no longer appropriate: reticular cells and endothelial cells lining blood vessels are quite distinct in their origin and functions.



Fig. 3.26 Section of mouse liver following an intravascular injection of colloidal carbon. The hepatic macrophages (Kupffer cells) are black because of the large amount of carbon which they have phagocytosed. $\times 300$.

Cells of the system

Macrophages may be recognised by the avidity with which they engulf particulate material of various kinds (Fig. 3.26) the firmness with which they adhere to a glass surface, both *in vivo* and *in vitro*, and their morphological differences from the other 'professional' phagocyte, the neutrophil polymorph.

Cells of the mononuclear phagocyte system are scattered widely throughout the body. In some sites they are normally inactive and inconspicuous but are capable, on stimulation, of enlargement, increased metabolism, and the active phagocytic role of the macrophage. Cells of the system include the following (van Furth *et al.*, 1975):

- (1) The Kupffer cells, which form part of the lining of the hepatic sinusoids; similar cells in the vascular sinusoids of the bone

marrow, spleen, adrenal cortex and adenohypophysis, and in the lymphatic sinuses of lymph nodes.

- (2) Cells in the spaces of the network formed by the reticular cells in the medulla of lymph nodes and red pulp of the spleen.
- (3) Cells on the surface of the serous cavities. These are particularly numerous in the omentum, where they are aggregated to form the 'milk spots'.
- (4) Alveolar macrophages lying free on the surface of, and also within, the alveolar walls.
- (5) Histiocytes in connective tissues, osteoclasts in bone, and microglial cells of central nervous tissue.
- (6) The monocytes of the blood and their precursors (monoblasts and promonocytes) in the bone marrow.

Kinetics of mononuclear phagocytes

In experimental animal studies, it has been shown that the monocytes are produced in the bone marrow, circulate in the blood for a few days, and are the precursors of the cells listed in 1-5 above, which may persist for months or possibly years. The fate of the mononuclear phagocytes is not known. There is some evidence that they may re-enter the blood and pass to the lungs, to be excreted via the bronchi.

The kinetics of macrophages in inflammatory lesions have been studied extensively by Spector and others (see Spector and Mariano, 1975). In various types of granulomatous inflammation, and in various tissues, it has been shown that most of the macrophages are provided by migration of monocytes. Their life-span in chronic inflammatory lesions has been found to vary depending on the causal agent. With relatively strong cytotoxic agents, macrophage turnover is rapid, a continuous supply from the blood being necessary to maintain the macrophage population of the lesion. With more bland agents, turnover is much slower (i.e. weeks), and mitosis of macrophages in the lesion may be almost sufficient to maintain the population. Curiously, mitosis of macrophages from inflammatory lesions is accompanied by a high incidence of chromosomal abnormalities which, by precluding further divisions, must limit their local proliferation.

Morphology of macrophages

Macrophages are motile cells and can assume polarity and various shapes. They have an oval, indented or irregular nucleus, abundant cytoplasm rich in lysosomes, and surface microvilli (Fig. 3.27). Phagocytic vacuoles, if present, are helpful in their recognition. Macrophages vary greatly in size (Fig. 3.28): they are usually larger than monocytes, but in the resting state they may closely resemble lymphocytes (e.g. in the peritoneum) and distinction can be made by histochemical demonstration of macrophage cytoplasmic enzymes.

Monocytes may be regarded as immature macrophage precursors which, on stimulation, transform to macrophages. The change involves increase in motility, size and phagocytic activity: cytoplasmic RNA and lysosomes increase, and the nucleus becomes larger and less

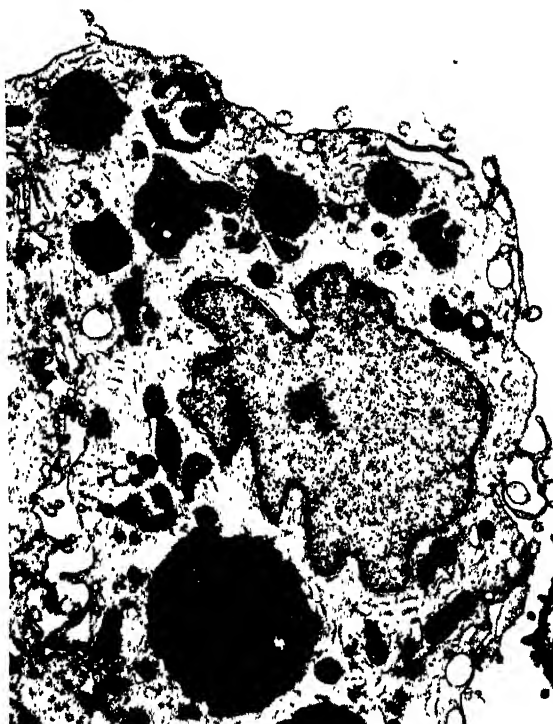


Fig. 3.27 An electron micrograph of part of a macrophage, showing microvilli (*top*) and a portion of the nucleus (*right centre*). The cytoplasmic dense bodies (black) are phagosomes formed by the fusion of lysosomes with phagocytic vacuoles and the 'empty' spaces are dilated endoplasmic reticulum. $\times 9000$.

condensed. Similar changes have recently been demonstrated by Wynne *et al.* (1975) when inflammatory exudate is added to cultures of macrophages (Fig. 3.29).

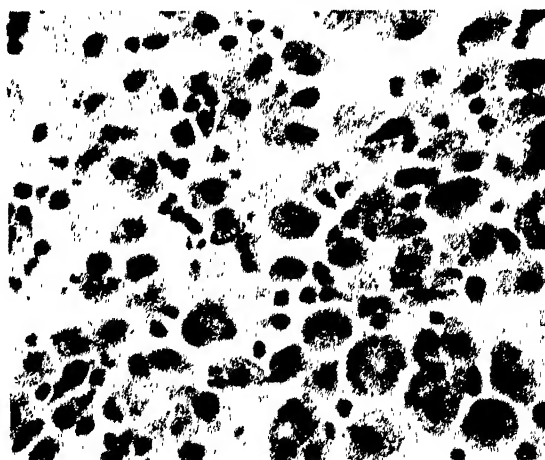


Fig. 3.28 Macrophages of various sizes in the wall of a chronically inflamed gallbladder. Note the ovoid or indented nucleus and abundant cytoplasm. $\times 405$.

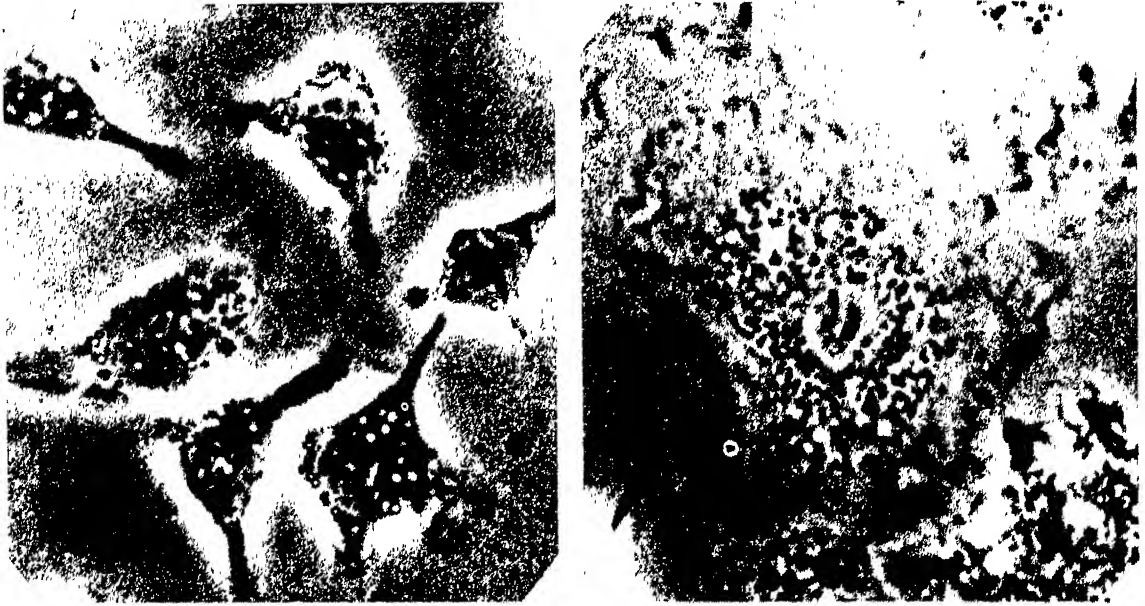


Fig. 3.29 Mouse peritoneal macrophages in culture, viewed by phase-contrast microscopy. Inflammatory exudate has been added to the culture shown on right. Note the increase in size, content of (phase-dense) lysosomes and (phase-lucent) vesicles and extensive cytoplasmic 'ruffling' of the stimulated cell. $\times 960$. (Professor W. G. Spector and Mrs. Katherine M. Wynne.)

In chronic inflammatory lesions, macrophages may assume the special features of **epithelioid cells** or of **giant cells**. Change to epithelioid cells (so-called because they have some resemblance to epithelial cells) involves an increase in the amount of cytoplasm and in rough endoplasmic reticulum: epithelioid cells are not actively phagocytic, and they may have a secretory function.

The factors responsible for change of macrophages to epithelioid cells are not fully understood: it is seen in immunological reactions of delayed hypersensitivity type, and may therefore be mediated by lymphoid cells: the same factors may be concerned in formation of Langhans' giant cells (see below) for these are usually associated with epithelioid cells.

Giant cells. It is common to see macrophages with two or more nuclei, but they can fuse together to form very large giant cells with sometimes over 100 nuclei. Such cells are usually classified into **foreign-body giant cells**, which have irregularly scattered nuclei, and **Langhans' giant cells**, in which the nuclei are arranged peripherally (Fig. 3.30). Spector has studied giant-cell formation by monocytes in tissue culture. At first they resemble foreign-body giant cells, but may later develop the features of Langhans' cells.

The **osteoclasts**, which digest bone matrix in the process of bone resorption, appear to be formed by coalescence of monocytes, and thus represent specialised cells within the macrophage system. The influence of the local environment on macrophages is also illustrated by the alveolar macrophages, which require oxygen for full phagocytic activity, whereas macrophages elsewhere can produce the necessary energy solely by glycolysis under anaerobic conditions.

Macrophage functions

The major function of the macrophage is phagocytosis and destruction of micro-organisms and other harmful or unwanted material. The general features of their scavenging role in inflammation are considered on p. 63 and their protective role in infection on pp. 180-81. Macrophages also have the important physiological role of removing dead or effete cells. For example they are responsible for taking up degenerate erythrocytes: they break down the cell and its haemoglobin, and release the iron, etc. for reutilisation.

Macrophages are capable of ingesting large amounts of insoluble material (Fig. 3.26), and of retaining it for months or even years. This



Fig. 3.30 Multinucleated giant cells formed by fusion of macrophages. The upper cell is a Langhans' giant cell in a tuberculous lesion: note the peripheral arrangement of the nuclei and abundant cytoplasm. The lower cell is a foreign-body giant cell: the nuclei vary in size and are irregularly distributed, and the cell has engulfed a fragment of suture material. $\times 750$.

happens when, for various reasons, abnormal amounts of lipids (p. 28) or iron accumulate in various tissues, or when dust particles are inhaled in atmospheric pollution. These conditions of abnormal storage are described in later chapters, in relation to the organs and tissues they most affect.

Other functions of the mononuclear phagocytes include the secretion of factors which stimulate fibrosis (p. 68), production of some of the components of complement (p. 142), participation in immune responses (p. 134) and the production of endogenous pyrogen in pyrexia (p. 189), and of colony stimulating factor which promotes proliferation and maturation of polymorph and monocyte precursors (p. 187).

Other cells in inflammation

Serosal cells. In acute inflammation of a serous surface, cells on or near the surface enlarge and often pass into the exudate at an early stage. They include macrophages and endothelial lining cells. In more chronic inflammation, the latter cells may grow in clumps in the fluid exudate in the cavity, and may show nuclear abnormalities and mucin secretion. In aspirated fluid, they are sometimes very difficult to distinguish from cancer cells.

Fibroblasts are seen in most acute inflammatory lesions, but their presence in large numbers is associated with chronic inflammation and repair. They are considered in the next chapter.

References and Further Reading

- Allison, A. C. (1978). Macrophage activation and nonspecific immunity. In *International Review of Experimental Pathology*, Vol. 18, pp. 304–346. Edited by G. W. Richter and M. A. Epstein. Academic Press, New York, San Francisco and London.
- Boyden, S. (1962). The chemotactic effect of mixtures of antibody and antigen on polymorphonuclear leukocytes. *Journal of Experimental Medicine* 115, 453–66.
- Burke, J. R. and Miles, A. A. (1958). The sequence, of vascular events in early infective inflammation. *Journal of Pathology and Bacteriology* 76, 1–19.
- Cochrane, C. G. *et al.* (1974). Soluble mediators of injury of the microvasculature: Hageman factor and the kinin forming, intrinsic clotting and fibrinolytic systems. *Microcirculation Research* 8, 112–21.
- †Cohnheim, J. (1889). *Lectures in General Pathology*, Vol. 1., pp. 242 to at least 270. (English

†Mainly of historical interest.

- translation). New Sydenham Society, London. (The classic account of the acute inflammatory reaction observed *in vivo*.)
- * van Furth *et al.* (1975). Mononuclear phagocytes in human pathology—proposal for an approach to improved classification. pp. 1–15. In *Mononuclear Phagocytes in Immunity, Infection and Pathology*, pp. 1062. Ed. by R. van Furth. Blackwell Scientific Publications, Oxford. (A comprehensive multi-author text.)
- Harris, H. (1953). Chemotaxis of granulocytes. *Journal of Pathology and Bacteriology* 66, 135–46.
- Hill, J. H. and Ward, P. A. (1971). The phlogistic role of C3 leukotactic fragments in myocardial infarction in rats. *Journal of Experimental Medicine* 133, 885–900.
- * Hurley, J. V. (1972). *Acute Inflammation*, pp. 137. Churchill Livingstone, Edinburgh. (A clear and interesting account suitable for the general reader.)
- † Landis, E. M. (1927). Micro-injection studies of capillary permeability. I. Factors in the production of capillary stasis. *American Journal of Physiology* 81, 124–42.
- * Lepow, I. H. and Ward, P. A. (Eds.) (1972) *Inflammation Mechanisms and Control*, pp. 388. Academic Press, New York and London. (Review articles on selected topics by some leading workers.)
- Lewis, E. and Turk, J. L. (1975). Comparison of the effect of various antisera and cobra venom factor on inflammatory reactions in guinea-pig skin. *Journal of Pathology* 115, 97–109.
- † Lewis, T. (1927). *The blood vessels in the human skin and their responses*, pp. 322. Shaw and Sons, London.
- Majno, G., Palade, G. E. and Schoeff, G. (1961). Studies in Inflammation. II. The site of action of histamine and serotonin along the vascular tree: a topographic study. *Journal of Biophysical and Biochemical Cytology* 11, 607–26.
- † Metchnikoff, E. (1905). *Immunity in Infective Diseases*. (English translation.) Cambridge University Press, London.
- Miles, A. A. and Wilhelm, D. L. (1955). Enzyme-like globulins from serum reproducing the vascular phenomena of inflammation. I. An activable permeability factor and its inhibitor in guinea-pig serum. *British Journal of Experimental Pathology* 36, 71–81.
- Pappenheimer, J. R., Renkin, E. M. and Borrero, L. M. (1951). Filtration, diffusion and molecular sieving through peripheral capillary membranes. A contribution to the pore theory of capillary permeability. *American Journal of Physiology* 167, 13–46.
- * Ryan, G. B. and Majno, G. (1977a). Acute inflammation. *American Journal of Pathology* 86, 185–276. (A detailed review with an extensive bibliography.)
- * Ryan, G. B. and Majno, G. (1977b). *Inflammation*, pp. 80. Upjohn Company, Kalamazoo. (A beautifully illustrated, clear account, of convenient length for the general reader.)
- Sevitt, S. (1958). Early and delayed oedema and increase in capillary permeability after burns of the skin. *Journal of Pathology and Bacteriology* 75, 27–37.
- Simionescu, N., Simionescu, M. and Palade, G. E. (1975). Permeability of muscle capillaries to small heme-peptides: evidence for the existence of patent transendothelial channels. *Journal of Cell Biology* 64, 586–607.
- Spector, W. G. and Mariano, M. (1975). Macrophage behaviour in experimental granulomas, pp. 927–42. In *Mononuclear Phagocytes in Immunity, Infection and Pathology*, pp. 1062. Ed. by R. van Furth. Blackwell Scientific Publications, Oxford. (A comprehensive multi-author text.)
- † Starling, E. H. (1896). On the absorption of fluids from the connective tissue spaces. *Journal of Physiology* 19, 312–26.
- * Wilkinson, P. C. (1974). *Chemotaxis and Inflammation*, pp. 214. Churchill Livingstone, Edinburgh. (An authoritative account of chemotaxis including original observations by the author.)
- Wilkinson, P. C., Russell, R. J. and Allan, R. B. (1977). Leucocytes and chemotaxis. *Agents and Actions*, suppl. 3, 61–70.
- * Willoughby, D. A., Coote, E. and Turk, J. L. (1969). Complement in acute inflammation. *Journal of Pathology and Bacteriology* 97, 295–305.
- * Willoughby, D. A., Giroud, J. P. and Velo, G. P. (Eds.) (1977). *Perspectives in Inflammation, Future Trends and Developments*, pp. 638. M.P.T. Press Ltd., Lancaster, England. (Report of an international conference with accounts by many leading workers.)
- Wynne, K. M., Spector, W. G. and Willoughby, D. A. (1975). Induction of DNA synthesis in rat macrophages *in vitro* by inflammatory exudate. *Nature* (Lond), 253, 636–7.
- * Zweifach, B. W. (Ed.) (1973–4). *The Inflammatory Process*, 3 vols. Academic Press, New York and London. (Accounts on most aspects of inflammation by leading workers.)

* General texts, reviews, and reports of symposia.

† Mainly of historical interest.

Healing (Repair) and Hypertrophy

Healing

Reaction of tissues to injury varies greatly in different species of animals and in different tissues. **Regeneration**, i.e. the replacement of a single type of parenchymatous cell by production of more cells of the same kind may be seen in man but different tissues vary in their regenerative capacity. A helpful guide to the expected reaction to damage of any tissue is given by the division of somatic cells into three types.

(a) **Labile cells** are those which under normal conditions continue to multiply throughout life and include epidermis, alimentary, respiratory and urinary tract epithelium, uterine endometrium and the haemopoietic bone marrow and lymphoid cells.

(b) **Stable cells** normally cease multiplication when growth ceases but retain mitotic ability

during adult life so that some regeneration of damaged tissues may occur. This group includes liver, pancreas, renal tubular epithelium, thyroid and adrenal cortex.

(c) **Permanent cells** lose their mitotic ability in infancy and the classic example of this group is the neuron.

In many instances healing of an organ or tissue occurs by regeneration, the cells lost being replaced by proliferative activity of those remaining. However, when the injury involves a cell type inherently incapable of this or when other factors, e.g. interruption of blood supply, prevents restoration, healing occurs by the **formation of a fibrous scar** the development of which is best illustrated by the healing of a wound of skin and subcutaneous tissue.

Healing of skin wounds

Healing by first intention (primary union)

Primary union occurs when uninfected surgical incisions and other clean wounds without loss of tissue are closed promptly, e.g. by sutures. (Fig. 4.1A) It is characterised by the formation of only minimal amounts of granulation tissue. It is a rapid process and contrasts with healing by **secondary intention** which occurs in an open wound, the edges of which are not brought together by sutures (Fig. 4.1B). Wound infection also prevents healing by first intention. The sequence of events in healing by first inten-

tion is as follows: blood clots between the wound edges and on the surface where it dries to form a protective scab. While removal of clot and dead tissue is occurring, firstly by the action of polymorphs and later of macrophages, there is a rapid spread of epithelium beneath the scab to bridge the wound surface within the first two days. Within 3 to 5 days capillaries and fibroblasts grow in beneath the epithelium, and collagen formed by the fibroblasts begins to bind the wound edges together by the end of the first week, reaching a maximum in two or three weeks. Thereafter wound strength slowly increases over many months.

The wound clot and its removal

When an incision is made in the skin and subcutaneous tissue, blood escaping from cut vessels clots on the wound surface and fills the gap between the wound edges which, in sutured wounds, is narrow. The fibrin in the blood clot acts temporarily as a glue which holds the cut surfaces together, while the dehydrated blood clot on the surface forms a scab which effectively seals the wound. Excess of blood clot deeper in the wound delays healing, for it greatly increases the risk of infection and if not evacuated, will only slowly be converted to fibrous tissue (see organisation, p. 98). During the first 24 hours, there is a mild inflammatory reaction at the wound edges with exudation of fluid and migration of polymorphs and later of monocytes and lymphocytes. Blood clot is digested by enzymes from disintegrated polymorphs and this is aided from about the third day by macrophages, derived mainly from blood monocytes, which ingest and digest any remaining fibrin, red cells and cellular debris, converting macromolecules into useful amino acids and sugar. The macrophages, which are the dominant cell by 72 hours after wounding, probably also play an important part, perhaps

along with platelets, in attracting fibroblast precursors into the wound and in stimulating their proliferation. They may also promote the formation of new blood vessels (see below). These changes represent the acute inflammatory (exudative) phase of response to injury and are usually mild unless infection supervenes.

Epithelial regeneration

The first tissue to bridge the incisional gap is the squamous epithelium of the epidermis. Within 24 hours and extending from 3–4 mm around the wound edge there is enlargement and flattening of the basal cells with loss of prominence of rete ridges. Two processes then contribute to the closure of the gap. Close to the cut edge, cells from the deeper part of the epithelium begin to slide over each other; they *migrate* out over the wound surface and become flattened to form a continuous advancing sheet. *Proliferation* also occurs, mainly among basal cells in the epidermis and pilosebaceous follicles adjacent to the wound. Mitosis is rarely seen in the migrating cells but occurs later in the new epithelium. While the

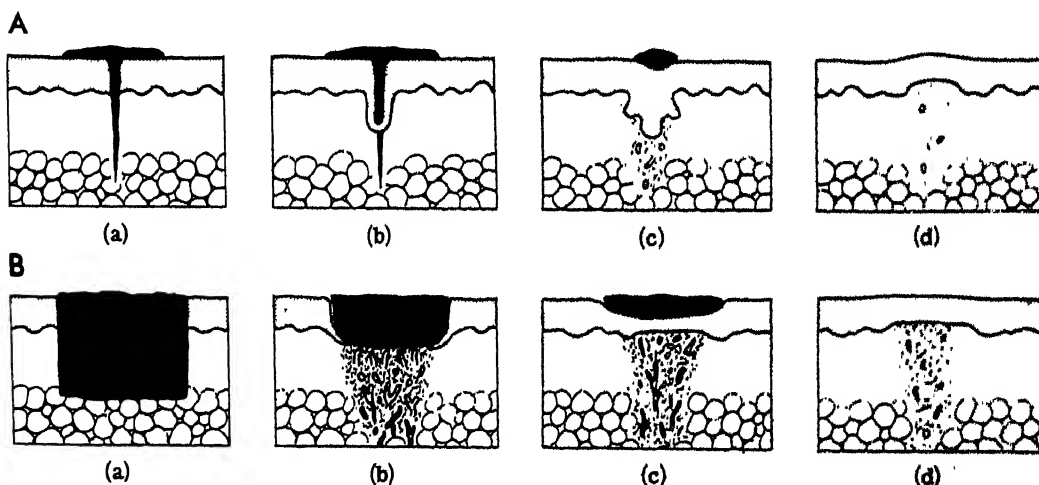


Fig. 4.1A. Healing by first intention (sutured surgical incision). Immediately after injury (a) the wound fills with blood clot and a scab forms on the surface. Epithelium which has grown from each side of the wound joins together within about two days (b) and later forms a spur of epithelium. A little granulation tissue grows into the wound (c). The collagen formed by the fibroblasts in the granulation tissue unites the wound edges. The epithelial spur resorbs and the epithelium does not re-form rete ridges. The narrow fibrous scar gradually becomes less vascular (d).

Fig. 4.1 B. Healing by second intention (open excised wound). After injury the wound fills with clot (a). Epithelium begins to extend under the clot and abundant granulation tissue grows into the base of the wound (b). Contraction of the wound occurs and the epithelium finally completely covers the base of the wound (c). The vascularity of this more bulky fibrous scar gradually diminishes (d).

advancing edge of the sheet of new epidermis consists of a single layer of flat cells, the older part at the periphery of the wound becomes stratified so that there is a gradient of thickness. The cells will only migrate over viable tissue. They burrow beneath the superficial part of the blood clot and wound debris, down the cut edges of the dermis; by their proteolytic enzymes, they cleave a path between dead and living collagen fibres (figs. 4.1A and 4.2). Within 48 hours the wound may be bridged by epithelium which rapidly becomes stratified but does not form rete ridges (Figs. 4.3, 4.8). Any epithelium which has grown down into the dermis is later resorbed (Fig. 4.1A).



Fig. 4.2 Aseptic abdominal wound showing the stage of healing at five days. The incision is represented merely by a vertical cellular line. The round body on the surface is a small scab beneath which the epithelium has extended down to cover the dermis. $\times 105$.

Suture tracks. Each suture track is a wound, with haemorrhage, death of cells and injury to skin appendages, and in consequence there is a slight inflammatory reaction and fibroblast proliferation (Fig. 4.4). Because the suture prevents closure of the surface epithelium the track is prone to infection and 'stitch abscesses' are commoner than sepsis of surgical incisions.



Fig. 4.3 Newly formed epithelium on healed ulcer. The epithelium is several cells thick, but there is little differentiation, and there is no formation of rete ridges. A similar appearance is seen in a healed surgical incision. $\times 400$.

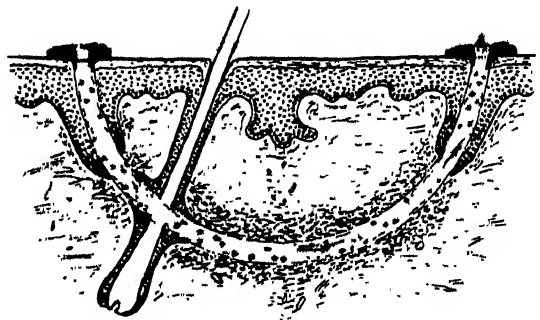


Fig. 4.4 Diagram of suture track a few days after wounding and before removal of sutures. In the centre of the picture the healed epithelium still forms a small projection into the dermis and beneath it the more cellular vertical line of the healing wound is seen.

The surface epithelium and that from a damaged hair follicle have grown along the suture track which is open to the skin surface and is lined by a layer of fibrin containing polymorphs. There is more vascular and fibroblastic proliferation around the suture track than around the original wound.

Epithelium also tends to grow down suture tracks from both ends (Fig. 4.4). Often much of the epithelium is avulsed when stitches are removed but some may remain and occasionally gives rise either to a small implantation cyst or to a subacute inflammatory reaction to keratin, which may simulate infection.

Repair of the dermis and subcutaneous tissue

These tissues heal by proliferation of new blood vessels and fibroblasts to form '**granulation tissue**', probably under the influence of stimulating factors released by macrophages. From about the third day, **vascular proliferation** is seen as capillary sprouts, which grow from blood vessels at the wound margins (Fig. 4.5), and advance up to 2 mm per day into the wound: the capillary sprouts are produced partly by rearrangement and migration of pre-existing endothelial cells and partly by their proliferation just behind the advancing tip. The

sprouts are often solid at first, but they unite with one another or join a capillary already carrying blood and develop a lumen. These newly formed capillaries are very delicate (Fig. 4.12), have an incomplete basement membrane and behave as if acutely inflamed: they leak protein-rich fluid with escape of some red cells, and polymorphs emigrate from them. It has been observed in rabbits that if blood flow is not soon established through a new vessel then the lumen disappears, the vessel reverting to a solid cord which then breaks and the ends retract by sliding back of endothelial cells to the nearest vessel carrying blood. Within a few days of the establishment of circulation, some of the new vessels differentiate into arterioles and venules by the acquisition of muscle cells either by migration from pre-existing larger blood channels or by differentiation from mesenchymal cells.

Lymphatic channels are re-established in the same manner as blood vessels.

Fibrous tissue production. After the removal of blood, fibrin and dead cells from the wound, and simultaneously with the development of new blood vessels, long, spindle-shaped fibroblasts (Fig. 4.6a and b) stream from the perivas-



Fig. 4.5 Capillary loops growing into a blood clot in a transparent chamber embedded in a rabbit's ear, photographed *in vivo*. Similar but less marked vascular proliferation is seen in the healing of a simple surgical incision. The rounded dark objects are macrophages which are digesting the blood clot in advance of growing vessels. (The late Prof. Lord Florey.)



Fig. 4.6a Fibroblasts in tissue culture. (The late Dr. Janet S. F. Niven.)



Fig. 4.6b Fibroblasts in a healing wound, showing the characteristic shape and early formation of collagen fibrils. $\times 350$.

cular connective tissue and begin to proliferate and to move into the wound. Within a few days fibroblasts thus come to lie in the wound: they produce both type I and type III collagen (see below), synthesis being greatest at about 7 days. The collagen fibres come to lie across the incision line and help to unite the cut edges from about the end of the first week after injury. Proteoglycan ground substance, also secreted by the fibroblast, may play an important role in determining fibre size and direction and also later in enhancing crosslinks between collagen molecules (see below). By the third week the total amount of collagen in the wound has almost reached a maximum. In contrast, at this stage the tensile strength of the wound is still low, but it increases over many months by further intermolecular bonding between collagen fibrils and by remodelling of the anatomical configuration of the collagen in response to mechanical stress. Such remodelling involves collagen turnover, i.e. synthesis and lysis which occurs also in normal (unwounded) tissues (see below).

While healing of a wound by fibrous tissue is clearly beneficial, in some tissues it may also have harmful results, for example narrowing of the oesophagus, stenosis of the mitral valve, or stricture of the urethra. Attempts to modify the fibrotic process by inhibiting collagen formation, promoting its destruction or altering its metabolism by drugs so far have not been very effective.

Collagen synthesis. Collagen is synthesised and secreted by fibroblasts in soluble form, and deposited extracellularly. As with secretory proteins in general, the polypeptide chains of collagen (**pro- α -chains**) are formed on the ribosomes with N and C terminal extension peptides. A distinctive feature of collagen synthesis is the conversion of proline and lysine residues on the growing polypeptide chains to hydroxyproline and hydroxylysine residues. This enzymic hydroxylation requires Fe^{++} , O_2 , ascorbic acid and α -ketoglutarate. Glycosylation of some of the hydroxylysine residues then occurs. In the cisternae of the endoplasmic reticulum pro- α -chains are converted to **pro-collagen** by the formation of disulphide bonds and they begin to assume a tri-helical structure. The pro-collagen molecules, of which there are several types, are stabilised by hydroxyproline and are secreted via the Golgi apparatus. Outside the cell the N and C extension peptides, which do not assume the helical form, are removed in some types of collagen by pro-collagen peptidase. This drastically alters the properties of the molecule which precipitates as **tropocollagen**. The tropocollagen molecules, which are rigid rods of $290 \times 1.4 \text{ nm}$, align side by side, probably in fives, staggered at a quarter of their length to produce a fibril with a 64 nm periodicity on electron microscopy (Fig. 4.7). The acquisition of tensile strength of the fibrils is



Fig. 4.7 Collagen fibres are seen here in longitudinal section. The characteristic, regular cross banding is evident. $\times 65\,000$.

Table 4.1 Types of collagen

Type	Molecular form	Tissue
I	Two identical chains $\alpha 1(I)_2$ and one $\alpha 2$	Dermis, tendon, bone, dentin, cornea (stains red with Van Gieson, green with Masson and blue with Mallory stains for collagen).
II	Three identical chains unique to itself $\alpha 1(II)_3$.	Cartilage.
III	Three identical chains unique to itself $\alpha 1(III)_3$	Embryonic dermis (and about 10–15% of adult) – early scar tissue, granulomas, cardiovascular tissue, synovial membrane (stains with silver nitrate as reticulin).
IV	Three identical chains unique to itself $\alpha 1(IV)_3$ but perhaps a heterogeneous group	Basement membrane (almost amorphous and does not show the typical banded structure on electron-microscopy).

dependent on the formation of co-valent links, mainly of aldemine and keto type. Increase in strength results from further alterations in the extent and nature of intermolecular crosslinkage.

Types of collagen. There are at least four different types of collagen derived from different structural genes (see Table 4.1). All have the triple helical configuration and types I, II and III have an identical appearance, banded at 64 nm, on electron microscopy. (Fig. 4.7).

The time needed for their synthesis and secretion varies as does their susceptibility to the various collagenases bringing about collagen lysis. While many tissues contain more than one type of collagen (e.g. adult dermis contains type I and about 10–15% of type III), mixed fibres are not found. Initially in a healing scar there are more type III (reticulin) fibres than in adjacent skin but later these are replaced by type I fibres.

Collagen lysis. Collagenases are formed, at the site where they are required, e.g. in healing wounds, by macrophages, polymorphs and regenerating epidermal epithelium. The collagenase is secreted directly onto the fibre by a closely apposed cell and this splits the fibre so that fragments may be ingested by macrophages. Splitting is more likely to occur in fibres with few or unstable cross links. Within the phagosomes, the fragments are broken down to amino acids or small peptides. In the healing wound, lysis occurs in the early stage of cleaning up the damaged collagen at the wound face and for a depth of about 7 mm into the surrounding tissue. There is also breakdown and replacement of much of the collagen first formed: wound remodelling continues for six months to a year. If the balance between synthesis and lysis is upset by starvation, sepsis, or deficiency of specific protein

or of oxygen, then the wound collagen may be extensively digested.

Events following primary wound healing

Once the wound has healed the young scar is commonly raised above the surface due to the underlying proliferative processes and is red as a result of increased vascularity. The blood vessels gradually decrease in number, probably in the manner already described, and the amount of collagen may also diminish. *Elastic fibres* are formed much later than collagen (Fig. 4.8). *Sensory nerves* may grow into the scar in about three weeks but specialised nerve endings such as Pacinian corpuscles do not re-form. The end result of healing by first intention should be a pale linear scar, level with the adjacent skin surface, but sometimes a **hypertrophic scar** or **keloid** forms (p. 342).



Fig. 4.8 Healed surgical wound of skin of 14 days' duration. The elastic fibres are stained black, and the healed wound is seen in the centre: it is composed of connective tissue in which elastic fibres have not yet formed. $\times 5$.

Healing by second intention (secondary union)

Healing of an open gaping wound with loss of tissue or of an infected closed wound occurs by the formation of granulation tissue which grows from the base of the wound to fill the defect. The vascular and fibroblastic proliferation which together make up the granulation tissue are much more abundant and healing takes much longer than when it occurs by first intention (Fig. 4.1 B). Skin grafts are increasingly used to speed the healing of open wounds.

Clean open wounds. As in the closed wound there is haemorrhage and exudation of fibrin from the cut surfaces. This is soon followed by a much greater emigration of polymorphs and subsequently of macrophages, from vessels in the wound: by enzymic action and phagocytosis these cells soften and remove the fibrin and other debris. As in the incised wound, epithelial cells at the margins enlarge and begin to migrate down the walls of the wound in the first day or two after injury. Migration and proliferation together produce a sheet of cells which advances in a series of tongue-like projections beneath any remaining blood clot or exudate on the wound surface. As the single layer of cells moves inwards towards the wound centre, there is stratification of the cells near to the wound margin (Fig. 4.9). Since the denuded

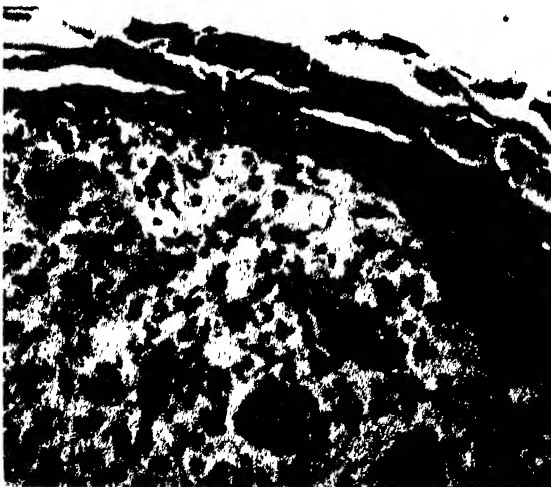


Fig. 4.9 Granulating wound with early growth of epithelium over the surface. The epithelium is growing from the right-hand side and tapers off as a thin layer. $\times 400$.

area is large, the advancing epithelial sheet does not completely cover the wound until the granulation tissue from the base (see below) has started to fill the wound space. Care should always be taken, in removing adherent dressings from an open wound, because the delicate epithelium is easily ripped off. As soon as the wound surface is covered, epithelial cell migration ceases and proliferation, stratification and keratinisation are rapidly completed, though rete ridges are not re-formed.

When the full thickness of the skin is destroyed by a burn re-epithelialisation is slow for the cells, with the help of their collagenases, have to burrow beneath the thick layer of dead coagulated collagen. In contrast, in a burn which destroys only part of the skin thickness, in a superficial wound or the donor site of a 'split thickness' skin graft, re-epithelialisation is relatively rapid as proliferation of epithelium occurs not only from the wound edge but also from the cut mouth of each pilosebaceous follicle. In man, slower and less perfect epithelial regeneration occurs from sweat gland ducts. If skin appendages are destroyed they are not re-formed.

Although epithelium shows the first evidence of reparative activity, within a few days the pre-existing vessels in the wound bed produce vascular sprouts which grow upwards, forming loops and coils near the wound surface (Fig. 4.10), giving it a red, granular appearance—hence the term '*granulation tissue*' (Fig. 4.11.) From these new, more permeable vessels (Fig. 4.12), small haemorrhages occur and polymorphs migrate, reinforcing those already present in the exudate on the wound surface and helping to keep down bacterial growth. At the same time as the new capillaries form, fibroblasts, some of which are in mitosis, are seen in the base and walls of the wound, often running parallel to the new capillary walls (i.e. towards the wound surface). This fibrovascular granulation tissue continues to proliferate and to fill the wound space only until epithelium grows over its surface, when the exudative inflammatory changes and the migration of polymorphs also subside. Later the fibroblasts may become orientated parallel to the wound surface (Fig. 4.13) and about the end of the first week collagen is produced and rapidly increases in amount. If epithelialisation is delayed, e.g. by further trauma or infection, granulations may

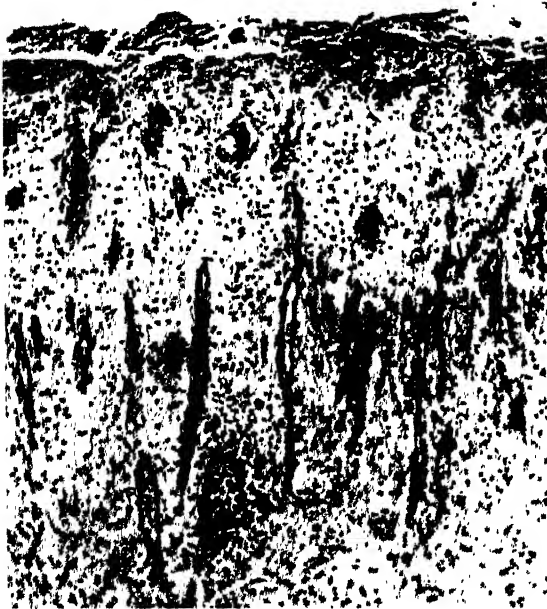


Fig. 4.10 Granulating wound showing the vertical lines of newly formed blood vessels. $\times 150$.



Fig. 4.11 Granulating wound of 12 days' duration. The advancing epithelial margin is seen as the dark surface layer on the right. Granulation tissue projects from the floor of the wound, and can be seen to contain many small blood vessels. $\times 10$.

pout from the wound surface. Following healing, there is gradual retraction and disappearance of some of the new vascular channels and further crosslinking and remodelling of collagen (see above) which, over a period of months, becomes progressively less cellular.

Healing of an open, excised wound is aided by contraction of the surface area in sites where the skin is mobile and loosely attached to underlying tissue. All edges of the wound do



Fig. 4.12 Newly formed thin-walled blood vessels in a granulating wound, the surface of which is to the top. $\times 150$.

not move together to the same extent, the degree of contraction being related to skin tension. This movement of the edges towards the centre of the wound is brought about by contraction of the fibroblasts which are now accordingly termed **myofibroblasts**. These cells develop temporarily a well-organised system of contractile cytoplasmic microfilaments resembling that of smooth muscle cells and responding to many pharmacological agents which contract smooth muscle. The bundles of parallel contractile microfilaments run longitudinally within the cells and are attached to the cell surface at sites where it is firmly adherent to other cells or to underlying tissue. Contraction of microfilament bundles in a meshwork of myofibroblasts in an open wound may thus pull together the wound edges, in which position they are stabilised by the deposition of collagen fibres. When desirable, as sometimes in flexor surfaces over joints, wound contraction may be inhibited by early skin grafting.

Infected wounds. The repair of infected wounds is accomplished by the same processes already described for clean, open wounds; that is, by the production of granulation tissue, but with a more pronounced acute inflammatory reaction, and also by formation of larger and more numerous blood vessels.

Open wounds, apart from those produced under aseptic surgical conditions, are almost



Fig. 4.13 Deeper part of granulating wound. *Below*, the collagen fibrils are being formed parallel to the surface; *above*, the vessels are seen running in a vertical direction. $\times 240$.

always contaminated by bacteria, and for this reason a careful surgical toilet should include removal of dead tissue, which promotes invasive bacterial growth; this *débridement* is an important part of treatment. Granulation tissue provides a good defence against bacteria because, being rich in small blood vessels, it can mount an effective inflammatory response. These local defensive factors may be aided by antibacterial therapy, which may permit early suture or skin grafting. If infection continues, however, more leukocytes pour into the surface exudate, which then becomes purulent. The result of acute infection on the healing processes is to inhibit both epithelial regeneration and proliferation of fibroblasts, so that healing is delayed. This delay and the greater tissue destruction result eventually in increased fibrous tissue and a larger and denser scar (Fig. 4.14).

Skin grafting. When a skin graft is placed on the recipient site it adheres to this new bed by fibrin and is nourished by diffusion of plasma from the raw surface. Within about three days

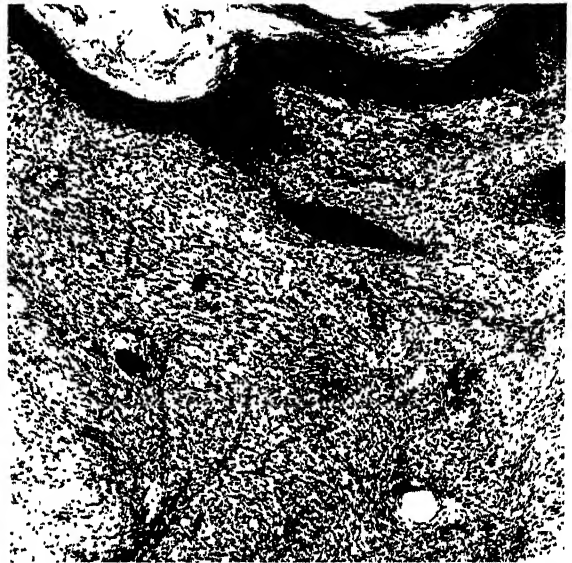


Fig. 4.14 Healing abdominal wound in which infection has delayed the process of healing. Note that the line of cellular tissue is much broader than in Fig. 4.2. $\times 75$.

capillary buds growing from the recipient area begin to unite with those on the undersurface of the graft, while ingrowing fibroblasts produce collagen which anchors the graft more securely. Good vascularity of the bed without haematoma formation, control of pathogenic infection and close, undisturbed contact with the graft, promote a rapid 'take'.

Grafts from the same individual (autografts) persist indefinitely. Grafts from another individual (homografts) are accepted initially but are destroyed within a month or so by an immune reaction by the recipient; they are nevertheless useful to provide temporary cover of extensive raw surfaces, for example large burns, and help to prevent excessive contraction of the area.

Control mechanisms in healing

While the morphological changes of wound healing are well known, the factors controlling the various observed processes are controversial or unknown.

Control of cell movement. The covering of a wound surface by epithelium cannot be explained simply by movement due to 'growth pressure'. While a burst of mitoses may displace adjacent cells by a concerted nudge, it is known that healing may occur without cell division and, in the skin, epithelial cell migration over the wound surfaces appears to precede proliferation. There is evidence from tissue culture that virtually all cells have the ability to

move along a surface to which they can adhere. If two cultures of fibroblasts are made on a plane surface the cells grow out as a monolayer from each explant until they collide, when movement virtually ceases because the cells will almost never pile up on one another. This is known as **contact inhibition**. The same phenomenon may govern the covering of the surface of a wound by epithelial cells from around the margin. The concept of contact inhibition in regard to epidermal cells and fibroblasts helps to explain not only initiation of movement but also the direction of cells into the wounded area, especially since there is no evidence that chemotaxis applies to any cell other than leukocytes (and spermatozoa). Contact inhibition appears to be tissue specific in some degree, for if skin and oesophageal epithelium meet, although both are squamous, cell movement and proliferation continue and cells heap up at the junction.

Explantation of a fragment of adult tissue into a culture medium does not result in rapid migration, although the cells at the free edge of the fragment now lack contact with similar cells. If the tissue is first wounded a short time before excision however, migration is greatly increased and it seems that some factor in addition to loss of contact inhibition may be involved. There is some evidence to suggest that there is a change in the cell surface, causing a diminution of its adhesive properties, thus permitting mobilisation.

The stimulus to migration and proliferation. It has been suggested that the stimulus to wound healing and enhanced mitotic activity is mediated by growth promoters—**trephones** or 'wound hormones'—liberated by damaged cells. Tissue culture studies suggest that there may be a stimulating substance but this has never been shown to initiate cell migration or multiplication *in vivo* and at the moment the existence of a wound hormone derived from damaged cells and promoting healing in animals remains no more than a possibility.

It has been postulated that cells of epidermis and other tissues normally secrete a diffusible tissue-specific depressor of cell mitosis and that wounding may inhibit the production or effect of this depressor substance or **chalone**, thus allowing an increase in mitotic activity. This possibility is supported by the experimental finding that removal of skin from one side of a mouse's ear provokes a burst of mitotic activity in the intact epidermis of the other side, maximal opposite the middle of the defect rather than opposite the wound edges. There is also evidence for the production of a fibroblast chalone in tissue cultures. Acceptance of the validity of this important concept of cell-specific chalones has not been universal because no one has yet purified and biochemically characterised these compounds, although their existence was reported many years ago.

The concept of the chalone has interesting implications apart from those related to wound healing. Their postulated inhibitory action on mitoses in skin cultures is apparently potentiated by adrenalin, and the known cyclical fluctuation in adrenal function may thus account for the diurnal mitotic rhythm seen in many organs. The adrenalin-chalone complex appears to act after DNA synthesis is complete, just before prophase in the mitotic cycle. There is some evidence also to suggest that tumour cells (which proliferate abnormally) may fail to synthesise or release adequate concentrations of tissue-specific chalones.

Factors which impair healing

Healing may be influenced by local factors. Infection delays healing (p. 85) as does a poor local blood supply; wounds of the relatively avascular shin tend to unite more slowly than those of the highly vascular scalp or face. Defects in collagen formation may result from generalised deficiency of vitamin C or of sulphur-containing amino acids and also from an excess of glucocorticosteroids.

Deficiency of vitamin C (ascorbic acid) and sulphur-containing amino acids. Man, monkey and guinea-pig are unable to synthesise vitamin C, and in the guinea-pig impaired synthesis of wound collagen occurs after even a few days on a diet lacking the vitamin. In man, however, a much longer dietary deficiency is necessary before collagen formation is depressed, although this may occur before scurvy is clinically apparent (p. 889). Patients with multiple injuries or extensive burns readily become deficient in vitamin C unless intake is increased. Deficiency of the vitamin disturbs the synthesis of collagen at the stage of intracellular hydroxylation of amino acids so that most of the underhydroxylated collagen remains within the cell and the small amount that escapes is more readily degraded than normal collagen. As a result, although wound contraction is not impaired, the wound is weak and tends to break down after re-epithelialisation and apparent healing: this was a well-known complication of naval surgery on scorbutic sailors. In addition to the reduced amount and abnormality of the collagen formed in the wound, the new capillaries may be unduly fragile because of failure of formation of basement membrane (type IV collagen). Deficient galactosamine may alter the properties of the ground substance. Similar alteration in collagen production with loss of wound strength may be seen in starving animals deficient in the sulphur-containing amino acids such as methionine, which are essential for collagen synthesis. Even when starvation continues, some of the methionine required for wound healing may be

obtained from endogenous tissue proteins but wound strength is increased by providing a diet adequate in protein. In well nourished individuals, protein and vitamin supplements will not speed healing or improve wound strength.

Excess of adrenal glucocorticoid hormones. Large doses of glucocorticoids, especially if given within the first few days after wounding, may suppress repair in experimental animals. In man the usual therapeutic doses seem to have little effect on healing of sutured incisions but may delay closure of open wounds with their higher energy requirements. Polymorphs and macrophages tend to be scanty, fibroblast proliferation and migration and the formation

of new blood vessels are all diminished, while epithelialisation and contraction are also deficient. In the steroid-treated patient there is a higher risk of wound infection and this is more likely to be undetected clinically. While the administration of vitamin A systemically or topically may counteract some steroid effects it does not restore wound contraction.

Zinc deficiency. Zinc is necessary for the synthesis of collagen. Oral supplements of zinc may promote wound healing in patients with zinc deficiency, but it is difficult to identify these patients since serum zinc levels may not reflect accurately the overall bodily status of zinc metabolism.

Healing of fractures

Healing in bones bears many resemblances to healing in soft tissues. Primary union of a fracture is however exceptional (p. 92), healing by the proliferation of callus (similar to wound healing by secondary intention) being the rule. There is initial haemorrhage and mild acute inflammation, followed by a proliferative or productive stage in which osteogenic cells play a vital part. Continuity between the bone fragments is first established by a mass of new bony trabeculae and sometimes cartilaginous tissue (**provisional callus**). This undergoes slow remodelling, with resorption and replacement so that, under favourable conditions, firm bony union is achieved. Sometimes restoration is so good that the fracture site is later hardly identifiable.

Early stages

A good deal of force is normally required to break a bone, the fragments are usually displaced, and in addition to a relatively small amount of haemorrhage between the bone ends much blood may seep into the tissues from ruptured vessels of the torn periosteum and adjacent soft tissues. In addition to haemorrhage, local inflammatory changes take place with hyperaemia and exudation of protein-rich fluid from which fibrin may be deposited. Polymorphs are scanty unless there is infection, and this is common only in **compound fractures**, i.e. when a bone fragment has torn the overlying skin or mucous membrane. Macrophages also

invade and phagocytose clot and tissue debris. Red blood cells are often removed rapidly from the fracture site, leaving a homogeneous mass of fibrin between the bone ends. A large amount of clot and debris between the bone fragments delays healing.

Bone necrosis occurs chiefly as a result of tearing of blood vessels in the medullary cavity, cortex and periosteum: the first recognisable histological evidence is observed within a day or two, the haemopoietic marrow cells showing loss of nuclear staining (Fig. 4.15). Fat released from dead marrow may be taken up by macrophages, and fat 'cysts' form, surrounded by foreign-body giant cells. Damage to the marrow may have serious results when globules of fat enter torn local venules and produce **fat emboli** in the pulmonary bed, brain and kidneys (p. 244). Because of its vascular arrangements the cortical bone usually suffers more extensive necrosis than the spongy medullary bone. The amount of bone necrosis depends especially on the local peculiarities of the blood supply; the talus, carpal scaphoid, and the femoral head following intracapsular fracture, are particularly liable to undergo extensive ischaemic necrosis. When there is splintering of bone (**comminuted fracture**) some of the fragments may lose their blood supply; they become necrotic and, if small, are eventually resorbed by osteoclasts. Bone death is recognisable histologically by loss of osteocytes from the bone lacunae (Fig. 4.15) but some cells may remain visible long after their death.



Fig. 4.15 The bone is necrotic and osteocytes have disappeared, leaving empty lacunae. Cell ghosts can be seen in the necrotic haemopoietic marrow. $\times 100$.

Provisional callus formation

(a) Periosteal reaction. The cells of the inner layer of the periosteum proliferate in a fairly wide zone overlying the cortex of each fractured bone end. (Fig. 4.16). A cuff of bone trabeculae is formed around each bone end at right angles to the cortex and anchored to it ((b) in Fig. 4.17).

Further woven bone trabeculae (p. 875), less well orientated, form an irregular meshwork whose pattern at this stage is uninfluenced by stress. This formation of new bone is dependent on the blood supply, which derives partly from surviving periosteal vessels but largely from muscle and other surrounding soft tissues. Mixed with this cuff of new bone there are often nodules of hyaline cartilage which usually do not appear until bone formation is well under way (Fig. 4.17). The amount of cartilage which is formed in provisional callus varies greatly from one species to another. Small mammals such as mice, rats and rabbits tend to form chiefly cartilaginous callus while in man the amount, though variable, is less. Cartilage formation is thought to be promoted by a poor blood supply and by shearing



Fig. 4.16 Rib 7 days after fracture. The fracture gap (a) contains fibrin and extends into the adjacent soft tissue. A spindle of highly cellular tissue (b) has formed in the muscle around the fracture site but only a little new subperiosteal bone (arrows) and cartilage have formed as yet. Bone and marrow at the fracture site are dead (c) but there is some early revascularisation of the marrow and a little bony callus is beginning to form in the medullary cavity (arrow). $\times 10$.



Fig. 4.17 Healing of displaced birth-fracture in the femur of a premature infant. The bone ends are cut obliquely. Bridging periosteal callus has formed but around the bone ends there is still a gap (a) which contains a meshwork of fibrin. The subperiosteal cuff of new bone is well seen at (b). The bridging callus consists partly of woven bone and partly of pale hyaline cartilage (c). The Haversian canals of the living cortex (d) are slightly enlarged by osteoclasts. $\times 5$.

strains and stresses so that it is particularly abundant in poorly immobilised fractures: bone gradually replaces the cartilage by endochondral ossification.

The two enlarging cuffs of callus advance towards each other and finally unite to bridge the fracture line leaving a gap between the bone ends (Fig. 4.17). This 'bandage' of **external callus** helps to immobilise the fragments in an unstable or poorly fixed fracture.

The amount of bridging periosteal callus varies greatly in different sites and under different circumstances. In intracapsular fractures (i.e. within a joint capsule), such as subcapital fracture of the femoral neck, the periosteum is lacking and union is almost wholly dependent on **internal callus** formed by osteo-

blasts lying in the medullary cavity (see below). By contrast, fractures of the shaft of large tubular bones, such as the femur and humerus, tend to form much external callus, internal callus in the relatively small medullary cavity not being striking. The formation of bulky external callus probably depends on plenty of surrounding undamaged muscle as a source of blood supply, for one of the causes of the difficulty in healing of fractures of the tibia is that they are partially covered by relatively avascular subcutaneous tissue and tend to form little callus. Poorly aligned fractures and those with much movement at the fracture site (e.g. the ribs and clavicle) are liable to produce abundant external callus, whereas fractures which are well immobilised by external or internal surgical fixation may unite with relatively little callus formation.



Fig. 4.18 Dead fatty marrow in the medullary cavity is being revascularised. A knot of proliferating capillaries is seen at (a). $\times 100$.



Fig. 4.19 A dead medullary bone trabecula is being removed by osteoclasts at (a) and new bone is being laid down on the surface of dead bone at (b). $\times 120$.

(b) Medullary reaction. The first evidence of healing is the advance of capillaries from the viable into the necrotic marrow (Fig. 4.18) closely followed by macrophages, fibroblasts and osteoblasts. The macrophages phagocytose and remove dead material, while osteoclasts begin to resorb dead spongy bone (Fig. 4.19) and the endosteal surface of the necrotic cortex. The osteoblasts produce new woven bone in the marrow spaces: the new bone is deposited partly on the surface of dead trabeculae which, when surrounded by new bone, may remain unresorbed for months or even years (Fig. 4.20).

In contrast with external callus, cartilage is rare in the medullary cavity, perhaps because it is a relatively vascular site and is protected from mechanical stress. It may form, however, when callus formation reaches the fracture gap (see below).

(c) Cortical reaction. The most striking reaction in the living cortex adjacent to the fracture is an increase in osteoclastic resorption with widening of the Haversian canals, presumably

partly due to disuse atrophy (Fig. 4.21 and p. 39). This may be followed later by some osteoblastic activity. Similar changes are seen in the dead cortex of the bone ends once there has been revascularisation of the Haversian canals from adjacent vessels in viable bone or from periosteal and medullary vessels. The resorption of necrotic bone may widen the fracture gap.

The fracture gap

The periosteal (external) callus unites the fragments externally but not directly across the bone ends. As already stated, immediately after fracture, blood clot, exuded fibrin and bony debris fill the gap and this is attacked by macrophages and by osteoclasts. The fibrin clot, which usually persists between the bone ends (Figs. 4.16 and 4.17), is finally invaded by blood vessels and cellular tissue containing varying amounts of osteogenic cells and fibroblasts, so that bony union may occur in either of two ways.



Fig. 4.20 Months after fracture dead bone trabeculae are still recognisable, covered by new living bone. $\times 70$.

(a) **Direct ossification** is brought about by osteogenic cells spreading from medullary and periosteal callus. Cartilage may also be formed and is converted into bone. This process is relatively rapid and effective.

(b) **Fibrous union** may occur initially. Fibrous tissue grows in from medulla or periosteum or both, becomes densely collagenised, and only then much more slowly becomes ossified (Fig. 4.21). This slower type of union occurs especially when there is instability, separation by excessive traction or marked resorption of the bone ends, massive necrosis, a poor blood supply, extensive periosteal damage, comminution or infection. Sometimes conversion to bone following fibrous union is very slow (**delayed union**) and occasionally it fails to occur (**non-union**). In non-union the fibrous tissue may become very dense, hyaline and finally fibro-



Fig. 4.21 This 12-week-old fracture of the distal fibula in an old man has formed a good deal of external callus medially (on the left). The fracture gap itself is filled with fibrous tissue in which a few trabeculae of metaplastic bone are forming. As a result of immobilisation and disuse the fibula is porotic with enlargement of Haversian canals. $\times 3$.

cartilaginous. The appearance of an area of eosinophilic fibrinoid necrosis is followed by a linear split which may enlarge and eventually develop a lining similar to synovium, thus forming a false joint (**pseudarthrosis**) (Fig. 4.22). The bone ends buried in the dense fibrous tissue tend to become very sclerotic.

Later stages: final remodelling

Once bony union has occurred and function has been regained, the bone begins to be remodelled in response to mechanical stresses. If the fracture has united at an angle new bone is incorporated on the concave side while

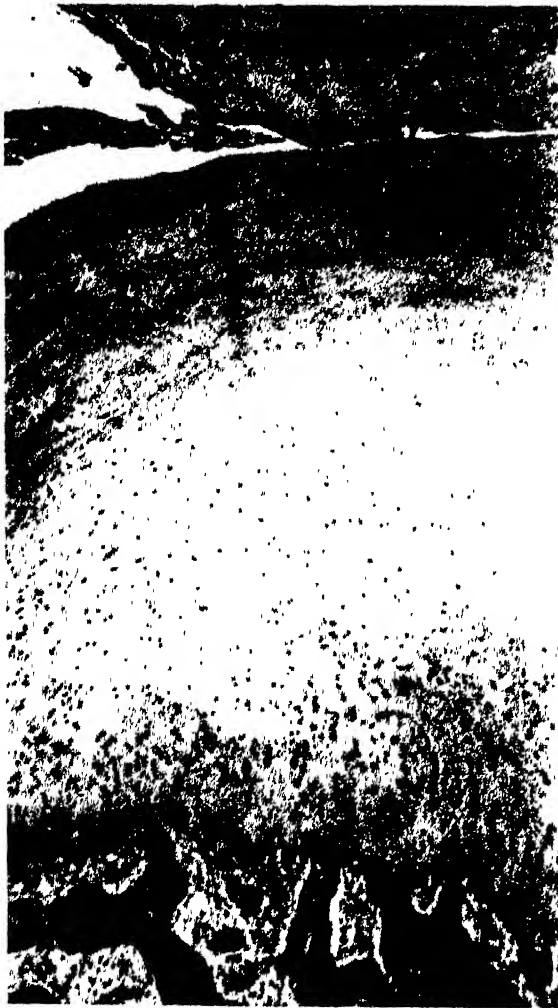


Fig. 4.22 Pseudarthrosis following fracture of the clavicle. The bone ends have become covered by cartilage. At the top of the picture a split in the cartilage has occurred giving a false joint. There is endochondral ossification of the proliferated cartilage in the lower part of the picture. $\times 40$.

resorption occurs on the convex, so that the bone becomes straighter. In any event excessive callus is resorbed, slowly formed lamellar bone begins to replace the hastily laid down woven bone, and any remaining necrotic bone is removed and replaced (Fig. 4.23). The cortex is re-formed across the fracture gap and gradually medullary callus is removed and the marrow cavity is restored. The whole process may take about a year and is more rapid and complete in children.

General factors and bone healing.

Some of the local factors affecting bone healing have already been mentioned but, as in wound healing, general factors are also important. Lack of vitamin C results in depression of both fibroblastic and osteogenic activity so that collagen and bone production are both deficient. Glucocorticosteroids administered to animals with fractures also delay healing but it seems that they have little effect when given to patients in the usual therapeutic doses. In vitamin D deficiency abundant callus may form, but it fails to calcify, remaining soft until the deficiency is made good.

Primary union of fractures

Although primary union of soft tissues is the rule in clean sutured surgical incisions, primary union in bone is a curiosity. It entails bony union with the formation of only minimal amounts of callus and was first described in compression arthrodesis (i.e. excision of the joint) of the knee, the cancellous surfaces of femur and tibia being held together by compression clamps. Bony union occurs in about 4 weeks and biopsy shows only a thin line of new bone at the contact points of opposing trabeculae. While moderate compression forces may assist union of a fracture, rigid fixation and close apposition of surfaces are probably the major factors. Cortical fractures in dogs, produced by a very fine saw with minimal necrosis, and fixed by a compression plate, have united without periosteal callus. The necrotic ends of the cortical bone were not resorbed but blood vessels entered the Haversian canals and subsequently the bone ends in contact were joined by new osteones (Haversian systems) which involved both bone fragments. At the opposite cortex from the compression device a small amount of new bone, formed from the endosteal cells of the Haversian canals, filled the narrow gap between the fragments and was later replaced by osteones. The healing process is similar to normal bone remodelling. Although a patient with a rigidly fixed fracture can be mobilised early, the more extensive bone necrosis results in slow healing since it depends on the formation of medullary callus and the direct replacement of cortical osteones.

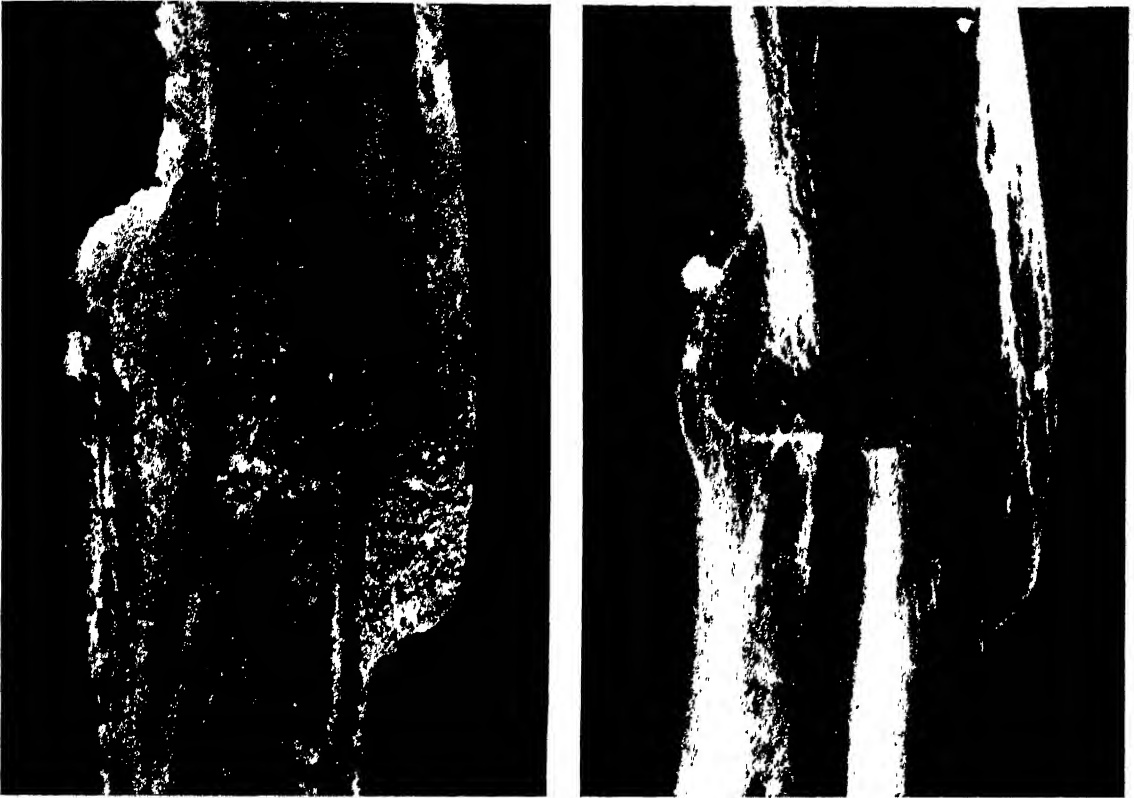


Fig. 4.23 This fracture through the midshaft of the femur united with anterior shift of the proximal fragment. Eighteen months after fracture there is firm bony union. The slab radiograph shows how the provisional callus has become remodelled along lines of stress with buttressing of the posterior and slightly concave part of the fracture line and some retubulation in the medullary canal at the fracture site.

Repair of some other tissues

Repair of articular cartilage

There is evidence that articular cartilage remains metabolically active throughout life, with continuous turnover of the proteoglycans and collagens of the matrix. Following injury, chondrocytes may proliferate to form cell clusters and there is an increase in proteoglycan turnover. However the capacity to form new collagen is limited and it is exceptional for these intrinsic reactions to produce any significant filling in of cartilage defects.

Extrinsic repair may occur by the growth of fibrous tissue over the articular surface from the joint margin. When there is loss of the full depth of the cartilage, fibrous tissue growing from the underlying marrow, through cracks in the exposed subchondral bone plate,

may cover the bone end. This collagenous tissue may then acquire a more chondroid matrix, to become fibrocartilage, and is sometimes able to function reasonably well. However, the amount of new tissue formed by extrinsic repair is usually insufficient and of poor quality, especially when the defect is large, so that joint function is only partly restored and tends to deteriorate.

Repair of tendon

A good functional result following healing of a severed tendon requires a strong fibrous union between the ends without loss of a full range of gliding motion. In patients with sutured tendons, repair occurs by ingrowth of fibroblasts and blood vessels from surrounding connective

tissue into the fibrin meshwork between the sutured ends. The cells, at first randomly arranged, later become orientated along the line of the tendon and produce collagen fibres. The amount of collagen synthesised is increased for some weeks after injury and, as in a healing wound, there is subsequent remodelling of the fibrous scar. The extrinsic source of vessels and cells during the healing process makes the formation of adhesions inevitable, but a good range of movement is retained when the adhesions are long and consist of loosely arranged areolar tissue. Restricted movement is associated with short adhesions containing large bundles of collagen fibres. Rough operative handling is thought to promote the formation of such adhesions, but little is known of other factors concerned.

Recent experimental evidence suggests that fibrocytes of tendon are not necessarily inert, but have the intrinsic potential for carrying out repair and remodelling without the formation of adhesions. It may be that it is the surgical suture of the severed ends which, by disturbing the local blood supply, impairs this response and makes healing dependent on extrinsic sources. Unfortunately suture is essential to hold the severed ends together and allow union to occur.

Repair of muscle

(a) **Skeletal muscle.** When muscle fibres are damaged, the sarcoplasm of dead fibres disintegrates, and this is followed by phagocytosis of the fragments by macrophages. Regeneration of muscle occurs in two ways. Firstly, by the formation of multinucleated sprouts (Fig. 4.24) from the surviving ends of injured fibres. It has been estimated that these sprouts advance into the damaged area at a rate of less than 1 mm/day. Secondly, by the growth of mononucleated myoblasts. The origin of these cells is controversial. Regeneration may be complete when the endomysial tube remains intact as in **Zenker's hyaline degeneration**. This condition may accompany severe toxic infections, especially typhoid fever, and tends to involve most severely the muscles of the abdominal wall, diaphragm and intercostals. When the connective and supporting tissue in muscle is also torn or destroyed, as in more severe injuries, regeneration is less well orientated and a

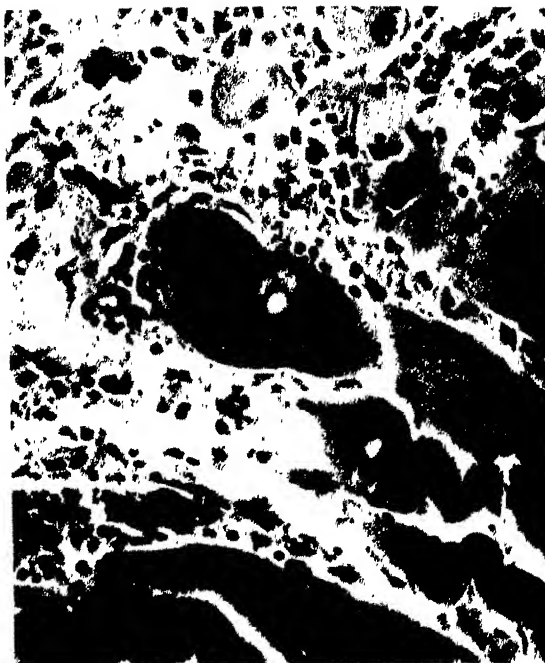


Fig. 4.24 Wounded skeletal muscle, showing sarco-plasmal sprouts with multiple nuclei. $\times 200$.

scar composed of fibrovascular tissue and irregularly orientated muscle fibres is formed. This is seen, for example, in Volkmann's ischaemic contracture (p. 930).

(b) **Visceral muscle (smooth muscle).** The healing of visceral muscle, e.g. in surgical incisions in the bowel or uterus, occurs by fibrous repair. Although smooth muscle may be seen in recently differentiated arterioles (p. 80) and in atheromatous plaques (p. 367), its origin is uncertain and it may arise by migration of smooth muscle cells or by differentiation from other mesenchymal cells. Proliferation with mitotic activity is said to occur in the early months of the physiological uterine enlargement of pregnancy.

(c) **Cardiac muscle.** Destruction of cardiac muscle by infarction is repaired by fibrous tissue (Fig. 15.11, p. 405). Effective regeneration occurs in young patients with Cocksackie virus infections or diphtheria, where there is damage to individual fibres with preservation of the endomysium, a situation similar to that of Zenker's degeneration in skeletal muscle.

Repair of nervous tissue and nerves

Central nervous tissue. Once mature nerve cells of the brain, cord or ganglia are destroyed,

they are not replaced by the proliferation of other nerve cells. There is also no useful regeneration of the axons in the central nervous system: indeed when an axon is severed at any point, the entire axon and the nerve cell body degenerate. Of the neuroglial cells, proliferation in response to tissue damage is restricted to astrocytes, this being referred to as gliosis (p. 728).

Regeneration of peripheral nerves. In contrast to the nerve cells and fibres in the central nervous system, peripheral nerves have considerable regenerative capacity.

When a nerve is transected, the axis cylinders distal to the cut undergo Wallerian degeneration, i.e. the axon and its myelin sheath break down and the debris is absorbed by macrophages. At the same time the Schwann cells proliferate within the neurilemmal sheath to form pathways along which the axons may regrow. Above the level of transection, myelin degeneration extends upwards only to the first or second node of Ranvier and the nerve cell body characteristically shows reversible changes (central chromatolysis p. 726). Axonal sprouts soon emerge from the proximal ends of the interrupted axis cylinders and, if the cut ends of the nerve are in close apposition, they grow into the distal part of the nerve and along the spaces formerly occupied by axis cylinders and now filled with proliferated Schwann cells. The axons grow at a rate of about 3 mm per day. These new axons, which at first are very thin, develop a new myelin sheath and then increase in diameter. They do, however, remain smaller than normal nerves unless they establish satisfactory end-organ connections and functional relationships. This takes some time, and restoration of function is accordingly slow and often imperfect.

A feature of regenerated nerves is that the internodal segments are shorter than in normal nerves.

The degree of functional recovery in a damaged peripheral nerve depends principally on the severity of the injury. If nerve fibres only are disrupted and the other components of the nerve trunk remain intact, as in a crush injury and sometimes in a stretching injury, regenerating axons can grow along their original endoneurial tubes, continuity of which is preserved and these fibres can re-establish their original end-organ relationships. If there is considerable

disorganisation of the internal structure of individual nerve bundles within the nerve trunk, the continuity of endoneurial tubes is less likely to be preserved: fibrosis then occurs within the damaged segment and interferes with the growth of axons from the proximal segment of the damaged nerve. Some axons never traverse the fibrous barrier, while those that do almost never grow along their original endoneurial tubes. Thus many axons fail to reach their original end-organ and abnormal and incomplete innervation results. When there is subtotal or complete loss of continuity of the nerve trunk, the proliferation of axonal sprouts, fibroblasts and Schwann cells from the proximal end of the nerve results in the formation of a so-called **traumatic or stump neuroma** (Fig. 4.25).



Fig. 4.25 Traumatic neuroma at severed proximal ends of nerves of an amputated arm.

In severely damaged nerves, surgical repair after excision of the involved segment or the traumatic neuroma is often the only hope of achieving any functional recovery, but some residual disability almost invariably persists.

Repair of mucosal surfaces

Cells which line mucosal surfaces, are being lost and replaced continuously throughout life and, like all surface epithelia, have a good potential for regeneration (Fig. 4.26). In general, the raw surface is first covered and only later is there differentiation into more specialised cells.



Fig. 4.26 Repair of lining of gallbladder after acute inflammatory desquamation. The epithelial cells extend as a thin flattened layer to reline the viscus. $\times 150$.

(a) **Gastro-intestinal tract mucosa.** Physiological replacement of lost surface cells takes place by proliferation of the cells of the mucosal glandular necks. Experimental excision of an area of mucosa, in the stomach for instance, is rapidly followed by re-epithelialisation: epithelial cells of the mucous neck type migrate over the exposed connective tissue, forming first a layer of thin, flattened epithelium, which later becomes cubical or columnar. The epithelium of glands adjacent to the wound undergoes mitosis, as do surface cells, and this proliferation keeps up the supply of migrating cells until the surface is covered. Gland crypts reform by mucous cells growing down into the underlying granulation tissue and, some weeks later, specialised cells, e.g. parietal cells, differentiate from the mucous cells in the crypts. Failure of re-epithelialisation of chronic peptic ulcers (p. 613) is not fully understood. Wounds of the mucous membranes following surgical anastomosis heal readily, and the line between the two different types of mucosa remains sharp.

Rectal lesions heal slowly with formation of much granulation tissue, but the small and, to a lesser extent, the large bowel mucosae have a good capacity for regeneration (Fig. 19.51d, p. 624). Repeated ulceration and repair of the colon, as for instance in ulcerative colitis and bilharzial infestation, may lead to overgrowth of the reparative mucosa, producing polypoid projections (Fig. 19.69, p. 638).

(b) **Respiratory tract mucosa.** The basal cells of the tracheal and bronchial lining epithelium proliferate throughout life and replace loss of the surface ciliated epithelium. Many microbial infections result in loss of only part of the thickness of the pseudostratified columnar epithelium, and this is readily replaced by cell proliferation. Destruction of the whole thickness of the mucosa is followed by the usual pattern of migration and proliferation of cells which at first appear transitional and later become low columnar: eventually the superficial cells develop cilia. Destruction of subepithelial tissue or repeated damage, e.g. in chronic bronchitis, may result in less perfect regeneration, the ciliated cells being replaced by columnar non-ciliated mucus-secreting cells, and under very unfavourable circumstances, as in heavy cigarette smoking, metaplasia (p. 441) of the regenerating epithelium to squamous type may be seen.

(c) **Urinary tract mucosa.** The urinary tract mucosa, like the epidermis, responds to injury by movement and proliferation of cells. The transitional cell epithelium of the bladder has particularly good powers of rapid regeneration, all layers of the mucosa participating.

Repair of kidney

The glomerulus is a highly specialised unit and little effective regeneration follows damage. Lost renal substance is replaced by fibrous tissue. If the basement membrane of the tubules remains intact, damage to tubular epithelium may be followed by proliferation (Fig. 4.27) and slow migration of surviving cells to restore continuity. It is doubtful, however, whether, when the damaged cells are highly specialised, as in the proximal convoluted tubule, the regenerated cells have the same degree of functional efficiency.

Repair of the liver

The hepatic parenchymal cells form a fairly stable population, and few mitotic figures are seen in the normal liver. Replacement of normal 'wear-and-tear' cell loss is by division of neighbouring cells. Some replacement is effected by division of the nucleus and enlargement of the cell, which explains the occurrence of binucleate and multinucleate liver cells, particularly in elderly individuals.

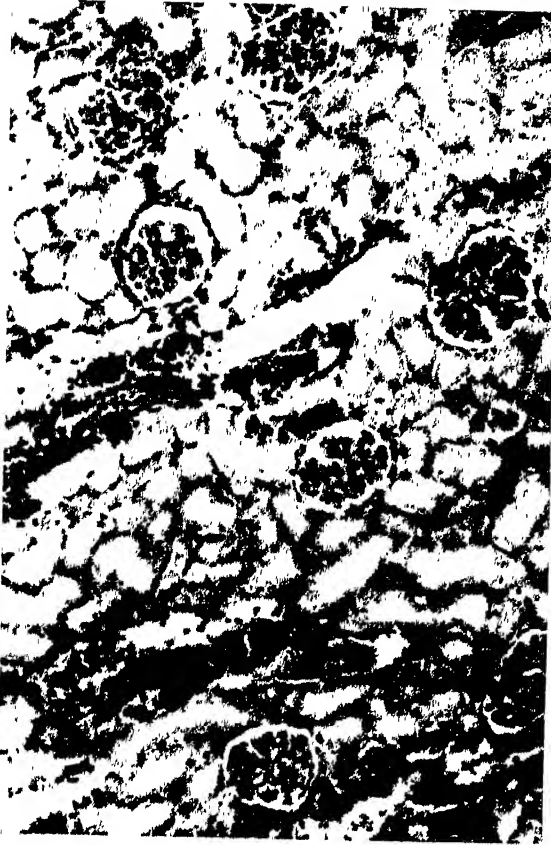


Fig. 4.27a Acute tubular necrosis in the rat. Note the anatomical normality of the glomeruli and the absence of nuclei in the dead tubular epithelial cells. $\times 150$.

Wounds of the liver are repaired by formation of connective tissue, with minimal replacement of parenchymal cells around the margin of the wound.

The outcome of liver cell necrosis depends on the distribution of the cells affected and also upon the presence or absence of an intact vascular system. Occlusion of hepatic arterial branches may be followed by infarction of liver tissue: as in most other tissues, coagulative necrosis occurs and the dead tissue is digested by macrophages and replaced by organisation (see below), eventually leaving a fibrous scar. Necrosis of individual liver cells scattered throughout the lobules, as in the typical attack of virus hepatitis, is followed by autolysis and disappearance of the dead cells, which are replaced by proliferation of surviving cells with restoration to normal. Even when there is destruction of all the parenchymal cells in the centres or mid-zones of the lobules the periportal cells proliferate and extend into the surviving vascular framework so



Fig. 4.27b Regeneration of renal tubular epithelium following acute tubular necrosis in the rat. Four mitotic figures are present. The adjacent regenerated cells are still of subnormal size. $\times 450$.

that normality is once again achieved. If, however, there is loss of most or all of the hepatocytes through the whole substance of the lobules, proliferation of the surviving hepatocytes leads only to irregular nodules of regeneration, the lobular pattern being lost. The vascular framework in the areas depleted of parenchymal cells becomes collapsed and scarred (Fig. 4.28).

The very high regenerative capacity of liver cells has been demonstrated by subjecting animals to excision of various amounts of liver tissue. After excision of two-thirds of the rat's liver, hypertrophy and hyperplasia in the remaining third result in restoration of a normal liver mass in 15–20 days. The capacity of the liver to regenerate in this way is maintained even when partial hepatectomy is performed monthly for up to one year.

The factors which regulate the extent of liver regeneration are not understood. When partial hepatectomy is performed upon one member of a pair of parabiotic rats, hepatocyte prolifera-



tion occurs in both animals. This and similar experiments suggest that a humoral mediating factor is responsible for hepatic regeneration. More recent studies in dogs suggest that a humoral growth-stimulating factor is responsible for hepatic regeneration and is released by regenerating liver. If the portal vein is divided and anastomosed to the inferior vena cava before partial hepatectomy, regeneration of parenchyma is impaired, and it thus seems that portal venous flow through the liver affects the degree of restoration. There is evidence that insulin plays a major role in maintaining normal functioning of liver cells and in promoting hepato-cellular proliferation. This makes good sense of the drainage of pancreatic venous blood to the liver and not to the systemic venous system.

Fig. 4.28 Following extensive liver necrosis there has been proliferation of surviving liver cells. These form the pale rounded nodules lying in a background of scar tissue from which dead liver cells have now disappeared.

Organisation

This means the replacement, by fibrous tissue, of solid, non-living material such as fibrin, clotted blood, intravascular thrombus and dead tissue. The process involves the gradual digestion of the material by macrophages. Since these phagocytic cells can only operate within a short distance of capillaries, removal of more than a small amount of dead material requires the ingrowth of capillaries and fibroblasts. This formation of granulation tissue is similar to that in healing, and the term organisation is used only when inanimate material is replaced by it.

An example already familiar is the removal of **fibrin deposited in acute inflammation** (p. 66). Thin strands of fibrin are rapidly phagocytosed and digested, but larger deposits, for example the thick layer which forms on the pleural surface in some cases of pleurisy, are removed more slowly by organisation. After the acute inflammation has subsided monocytes continue to emigrate from the vessels in the pleura underlying the adherent fibrin and

assume the features of macrophages: they begin to digest the fibrin by a combination of phagocytosis and release of lysosomal enzymes. Capillary sprouts develop from the superficial pleural vessels and grow into the spaces created by the macrophages (Fig. 4.29): they anastomose to form a network of capillaries, some of which enlarge and develop into arterioles and venules. The capillaries are accompanied by proliferating fibroblasts which produce collagen and ground substance. A thin layer of granulation tissue thus takes the place of the deepest part of the fibrin and gradually the process extends until all the fibrin has been replaced. Meanwhile, the granulation tissue slowly changes to less vascular, firm fibrous tissue. If the layer of fibrin has glued together the visceral and parietal pleura, organisation proceeds from both surfaces and meets in the middle: in consequence, the lung becomes firmly bound to the chest wall by fibrous tissue.

A second example of organisation is seen when haemorrhage occurs into the tissues and

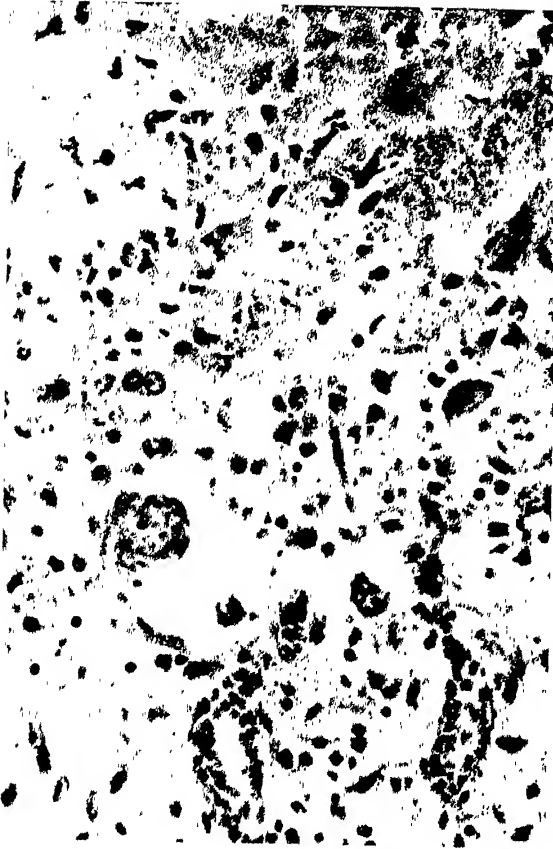


Fig. 4.29 Dilated capillaries and fibroblasts are beginning to grow into the dense fibrin on the surface of the lung. $\times 400$.

the escaped blood clots to form a solid mass, i.e. a **haematoma**. This is removed by organisation from the surrounding tissues (Fig. 4.30) and progresses to the centre of the clot, even-



Fig. 4.30 Organisation of a haematoma after 7 days. Capillary sprouts are beginning to grow into the clot and there are also large numbers of macrophages at its margin. $\times 80$.

tually leaving a fibrous scar. A patch of **necrotic tissue** (most commonly seen in the heart, brain, kidneys, etc. as the result of arterial blockage by thrombus, i.e. an infarct) is similarly removed by organisation with consequent scar formation.

When an artery or vein is blocked by **thrombus**, this is removed partly by organisation, but other processes are involved and the fate of thrombi is described on pp. 241–3.

Hypertrophy

Stimulation of the parenchymal cells of an organ, usually by increased functional demand or by hormones, results in an increase in the total mass of the parenchymal cells. This may be brought about by enlargement of the cells—**hypertrophy** or by an increase in their number—**hyperplasia**. The relative importance of the two processes varies in different organs. In some, e.g. the skeletal muscles, enlargement is purely by hypertrophy, but in most organs hypertrophy and hyperplasia both contribute.

(a) The response to increased functional demand

This is illustrated by the hypertrophied muscles of manual labourers and athletes; the individual fibres increase in thickness but not in number. Similarly, when extra work is demanded of the heart as a result of valvular disease or high blood pressure (Fig. 4.31), there may be much thickening of the muscular walls of those chambers which bear the brunt of the extra work. Narrowing of the mitral valve, for example, produces chiefly left atrial and right



Fig. 4.31a Hypertrophied muscle fibres of heart in a case of arteriosclerosis with high blood pressure. $\times 250$.

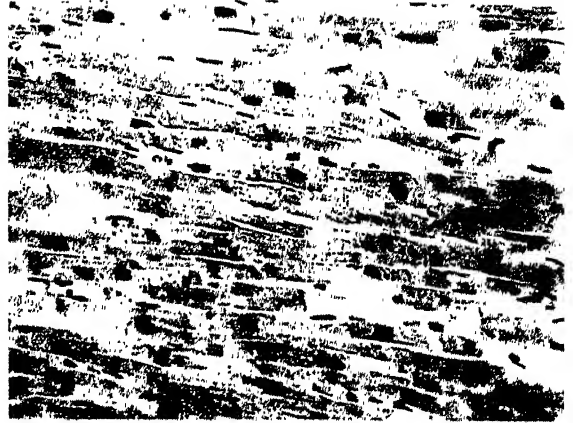


Fig. 4.31b Slightly atrophied heart muscle; to compare with Fig. 4.31a. $\times 250$.

ventricular hypertrophy, whereas systemic hypertension gives rise predominantly to left ventricular hypertrophy (Fig. 15.1, p. 399). The heart may increase to twice the normal weight, the degree of hypertrophy being limited by the diffusion of oxygen and nutrients between the capillaries and the thickened fibres. **Smooth muscle** may also undergo hypertrophy, for example in the wall of the stomach in a patient with pyloric stenosis, in large bowel proximal to an obstructing tumour (Fig. 19.81, p. 653), or in the bladder when the outflow is narrowed by an enlarged prostate (Fig. 25.3, p. 992). The muscle in arterial walls also hypertrophies in response to long continued high blood pressure (Fig. 14.13, p. 371). The most striking hypertrophy is seen in the pregnant uterus, where a combination of increased functional demand and hormonal stimuli results in enlargement of fibres to more than a hundred times their original volume. In early pregnancy, there may be both hypertrophy and hyperplasia of muscle fibres. After parturition, the muscle fibres return to a normal size and this is seen also in hypertrophied heart muscle when the increased work stimulus is removed.

In some instances, the response to an increased demand is by a pure hyperplasia; for instance, blood loss is not followed by enlargement of red cells and leukocytes but by an increase in their production in the haemopoietic marrow (Fig. 17.9b, p. 517).

Compensatory hypertrophy may occur in the survivor of a pair of organs when one is removed. Following nephrectomy, the remaining

kidney enlarges and, particularly in young patients, may double its weight. This is brought about by increase in the size of the nephron as a result of both hypertrophy and hyperplasia of the component cells of the glomeruli and tubules. Removal of one adrenal leads to hypertrophy of the cells of the opposite cortex, the medulla remaining unchanged. Following removal of a lung, the remaining lung enlarges but this is caused mainly by over-distension, which produces a lasting enlargement of the alveoli; only when it occurs in early life is there any formation of new alveoli. Compensatory hypertrophy is a widespread natural phenomenon, as illustrated in Fig. 4.32.

The testes are exceptional in that, in both man and animals, removal of one in adult life is not followed by enlargement of the other, the number of spermatozoa produced being reduced.

(b) Hypertrophy due to hormonal changes

A balanced activity of certain of the endocrine glands is essential for the normal growth and metabolism of the tissues and many of the examples of hypertrophy and hyperplasia already mentioned require the continued physiological action of the growth hormone of the anterior pituitary as well as an adequate blood supply. Excessive secretion of growth hormone (usually due to a tumour of the oxyphil cells of the anterior pituitary) results in adults in acromegaly (p. 1010) with bone enlargement and generalised organ and tissue hypertrophy. This is



Fig. 4.32 This beech tree was largely uprooted by a gale six years before, and has since leant against its neighbour. It continues to survive because a few roots (on the right) have retained contact with the soil and have become greatly hypertrophied.



Fig. 4.33 Breast lobule in pregnancy showing marked hypertrophy and hyperplasia. $\times 50$.

even more strikingly seen when the excess of hormone occurs in adolescence, before the skeleton matures; the result is gigantism (p. 1011). An example of physiological hormonal hyper-

trophy is the enlargement of the breasts in pregnancy, when the formation of mammary gland acini is stimulated chiefly by hormones from the corpus luteum or placenta (Fig. 4.33).

Further Reading

- Guber, S. and Rudolph, R. (1978). The myofibroblast. *Surgery, Gynecology and Obstetrics* **146**, 641.
- Hunt, T. K. and Van Winkle, W. Jr. (1976). *Fundamentals of Wound Management in Surgery. Wound Healing: Normal Repair*. Distributed by Smith Kline and French Laboratories.
- Jackson, D. S. (1978). Collagens. In *Diseases of Connective Tissue*, pp. 44–8. Ed. by D. L. Gardner. Publ. by *Journal of Clinical Pathology*.
- Matthews, P. and Richards, H. (1974). The repair potential of digital flexor tendons. An experimental study. *Journal of Bone and Joint Surgery* **56B**, 618.
- McKibbin, B. (1978). The biology of fracture healing in long bones. *Journal of Bone and Joint Surgery* **60B**, 150.
- Sevitt, S. (1970). Bone repair and fracture healing. *British Journal of Hospital Medicine* **3**, 693.

Immunophysiology: The Immune Response

Introduction

The invasion of the body by living organisms, including viruses, bacteria, and protozoan and metazoan parasites, presents a major threat to the stability of the internal milieu upon which Claude Bernard placed such importance. To counter this threat certain general defence mechanisms have evolved—a relatively impermeable epidermis, methods of ridding the body of noxious material, such as vomiting, diarrhoea and coughing, the dilution of irritants by increased flow of interstitial fluid in inflammatory oedema, and the destruction of particulate matter by phagocytic cells. In addition, there exists in vertebrates a special defence mechanism of immense potential which is mobilised when the body is invaded by foreign organisms and which is expressly and specifically adapted to overcome the effects of any particular invader. The special mechanism is called **acquired specific immunity** and its study—the science of **immunology**—is of great importance in the understanding and prevention of disease.

The phenomenon of acquired specific immunity has been recognised for centuries in that individuals who had survived an attack of certain clearly recognisable infectious diseases such as smallpox were known to be much less susceptible to the disease during a later epidemic. Such individuals could be said to show **immunity** (i.e. protection) against the disease, **acquired** inasmuch as it did not apparently exist before the first infection, and **specific** inasmuch as an attack of smallpox protected the individual against a further attack of smallpox but had no bearing on his susceptibility to later attacks of measles, diphtheria, etc.

This knowledge has been applied with great success to the prevention of infectious disease by prophylactic immunisation, a procedure in which a relatively harmless variant, or modified toxin, of a pathogenic organism is purposely

introduced into the body; this results in the development of specific immunity such as would be encountered following recovery from the natural disease. The principle is well illustrated by Edward Jenner's use of fluid from the lesions of cowpox (vaccinia) to vaccinate against smallpox; it was known to Jenner that milkmaids who had had natural cowpox infection had developed not only markedly altered reactivity to re-infection with cowpox but also resistance to a first infection by smallpox, a closely related but much more serious disease. In this case the two viruses are so similar that immunity to one is effective also against the other.

When an individual has become immune following natural infection or prophylactic exposure to a pathogenic organism or its toxin he is said to be **actively immunised** against that organism. Specific resistance to infection can in many instances be conferred upon a non-immune individual by an alternative method, namely the injection of *serum* or *lymphoid cells* from an actively immune individual. The state of **passive immunity** so conferred is not due to transfer of the infecting organism or its toxin but to the transfer of the products of immunisation developed by the actively immunised donor of the serum or lymphoid cells. The *passive transfer* of immunity thus provides a way of analysing the factors which contribute to the immune state, and by transfer experiments it has been shown that in some instances specific immunity results from the presence in the serum of special globulins known as **antibodies**, while in other cases the immune state seems to be mediated directly by **specifically primed (sensitised) lymphocytes** without the participation of serum antibody. Serum containing one or more antibodies produced by active immunisation is termed **antiserum** or **immune serum**.

There is obviously considerable survival advantage in having the ability to acquire specific

immunity to pathogenic organisms or their toxins. Unfortunately the specifically altered reactivity produced by an immune response can lead to reactions which result in tissue injury. This can happen following exposure even to relatively harmless substances, such as grass pollen, which on subsequent contact causes a harmful and seemingly unnecessary inflammatory reaction in certain individuals. Such **acquired specific hypersensitivity** has in many cases been shown to be due to reactions of an immunological nature although with its connotation of protection the word 'immune' appears unsuitable. Because of this difficulty and the gradual way in which knowledge of the complex processes involved has unfolded, an elaborate but imprecise jargon relating to immuno-

logy has developed and many authors have used the same terms with different meanings. Von Pirquet, for example, coined the word **allergy** as a unifying term to indicate *altered* specific reactivity of all kinds, including both protective immune responses and also hypersensitivity. Despite this the words allergy and hypersensitivity are frequently used interchangeably.

In this book we follow the current common practice of using the words 'immune' and 'immunity' in two distinct ways: in one they are general terms to embrace all forms of specifically altered reactivity (i.e. allergy in von Pirquet's sense) and in the other they refer to specific protection against disease; we believe that the meaning implied will be evident from the context.

Antigens

An antigen is a substance which is **immunogenic**, i.e. capable of evoking an immune response: this may take two main forms.

(1) **Antibody production**, i.e. the appearance of globulin molecules which have the property of combining specifically with and remaining attached to antigen of the same kind as that which has led to their formation.

(2) **Cell-mediated immunity**, i.e. the production of specifically primed lymphocytes whose presence can be demonstrated *in vivo* by the development of a local inflammatory reaction appearing about 24 hours after intradermal injection of the antigen—a **delayed hypersensitivity reaction**.

Most antigenic stimuli evoke both cell-mediated immunity and antibody production. These responses take place in the lymphoid tissues, and their products, specifically primed lymphocytes and antibody, both capable of reacting with the antigen, are released into the bloodstream.

Under certain conditions, an antigenic stimulus may induce a state of **specific immunological tolerance**, i.e. non-responsiveness to subsequent challenge with the same antigen.

The term '**immunological reaction**' should not be confused with '**immune response**' as described above. An **immunological reaction** is the effect observed when the products of the immune response—antibody or primed lym-

phocytes—encounter and combine with the appropriate antigen *in vitro* or *in vivo*.

Factors affecting the immune response

The form taken by the immune response depends upon several factors including the nature of the antigen, the genetic constitution of the individual exposed to the antigen, the route by which the antigen enters the body and the dose administered. These factors are discussed below.

The nature of antigens. It is difficult to define precisely the properties which make a substance capable of evoking an immune response, but in general terms antigens are large molecules, usually of molecular weight exceeding 3000, fairly rigid in structure, and either protein or carbohydrate, with or without associated substances such as lipids. Antigen-antibody reactions appear to be the result largely of stereochemical interactions of molecules of complementary configurations, analogous to the interaction of lock and key. For this reason floppy molecules such as gelatin are poor antigens. Evidence will be presented later that immune responses follow the binding of antigen molecules to specific receptors on the surface of lymphocytes. It seems likely that, to trigger off an immune response, antigen must form a link

between these surface receptors: this explains why most antigens are large molecules.

Although some small molecules, such as *p*-**para**-aminobenzoic acid, are not by themselves antigenic, they may become so if they are attached to larger molecules. Injection of *p*-aminobenzoic acid attached by diazotisation to serum albumin may result in formation of some antibody molecules which combine specifically with the *p*-aminobenzoic acid moiety and not with the albumin carrier protein. In these circumstances, *p*-aminobenzoic acid is said to be a **hapten**, i.e. a substance which is antigenic inasmuch as it can take part in an immunological reaction (in this case antigen-antibody combination) but which is not itself immunogenic, i.e. cannot by itself evoke an immune response (in this example, the formation of specific antibody) unless it is conjugated with macromolecular material. The existence of such simple haptens suggests that the antigenic specificity of large molecules may be determined by the three-dimensional configuration of small parts of these molecules (**antigenic determinant sites** or **epitopes**). Study of synthetic polypeptide and polysaccharide antigens has confirmed that specific antigenic determinant sites do consist of a few amino acids or monosaccharides, and it has been shown that most naturally-occurring macromolecules such as plasma albumin contain several antigenic determinant sites of differing specificity. Larger, complex antigenic particles, such as bacteria, contain a correspondingly greater number and variety of epitopes. Both antibody production and cell-mediated immunity evoked by antigenic material are correspondingly complex, each particular kind of epitope being potentially capable of inducing both types of response.

Genetic constitution of the individual. The repertoire of specific immune responses which an individual can mount depends on the selection of genes inherited from two distinct sets—the V genes which code for the various antigen binding sites found on different antibody molecules (pp. 126–8) and the Ir (immune reactivity) genes which are not linked with (i.e. on the same chromosome as) the genes coding for antibodies, but lie in the 'I' region of the major histocompatibility complex (p. 167). The genetic nature of the responsiveness is clearly shown by crossing inbred strains of mice which exhibit marked differences in their responses to im-

munisation with substances containing only one type of epitope, such as simple synthetic oligo-peptides. Inheritance of specific responsiveness to most natural antigens is difficult to demonstrate because of the multiplicity and variety of antigenic sites on natural macromolecules but the magnitude of response to these substances is controlled non-specifically by a number of genes, some affecting such characters as rate of antigen degradation by macrophages (pp. 134–6).

Genetic factors also play a large part in determining the antigenicity of tissues, and this field has become particularly important in the practice of blood transfusion and of tissue transplantation. The molecular composition of tissues, including potential antigenic sites, is, of course, genetically determined. In general, when tissues are injected or transplanted from one individual to another, the more genetically dissimilar or foreign the two individuals are to each other, the easier it is to induce antibody formation and cell-mediated immunity. For example, human red cells injected into rabbits evoke a wide variety of antibodies reacting with a corresponding variety of antigenic determinants on the human red cell; when, as in this case, the antigen is derived from a species other than that of the immunised animal (and this would include bacteria, etc.), it is called a **hetero-antigen** and the antibodies are **hetero-antibodies**. Injection or transplantation of one human with the red cells or tissue cells of another may result in the formation of antibodies to antigenic groups not shared by both individuals; e.g. human red cells containing the Rhesus antigen D (Rhesus positive cells) into a person whose cells do not contain this antigen (Rhesus negative) may result in the development of antibodies specific for D antigen; antigens which differ within a species are called **iso-antigens** and the corresponding antibodies **iso-antibodies**. In general, iso-antigens are much less numerous and less likely to evoke an immune response than hetero-antigens. Finally it should be noted that injection of an individual with his own cells (**auto-antigen**) results in **auto-antibody** formation or cell-mediated immunity only in exceptional cases; the subject of **auto-immunity** is considered further on p. 161. Unfortunately the prefixes used to indicate the relationship between individuals providing antigen and forming antibody are, by usage,



Table 5.1 Terminology of antigens, antibodies and tissue grafts

Relationship between donor and recipient	Genetic terminology	Antibody, antigen	Transplantation terminology
Same animal	—	Auto-antibody Auto-antigen	Autograft
Identical twins and inbred strain	Syngeneic (Isogeneic)	—	Isograft
Same outbred species or different inbred strains	Allogeneic	Iso-antibody Iso-antigen	Allograft (Homograft)
Different species	Heterogeneic Xenogeneic	Hetero-antibody Hetero-antigen	Xenograft (Heterograft)

616-02
AND

different from those used in the more recent field of tissue transplantation, in which graft rejection is effected mainly by delayed hypersensitivity reactions. Table 5.1 summarises this confusing and irrational situation.

As stated above, the cells and tissues of an individual are antigenic when injected or grafted into an animal of another species or even into a different individual of the same species, yet with certain exceptions the individual does not react to the antigens of his own cells by the development of auto-antibody or delayed auto-hypersensitivity. Non-reactivity to auto-antigen is a general physiological principle described by Ehrlich as 'horror autotoxicus'. It may reflect absence of lymphoid cells with the genetic coding necessary for the synthetic processes associated with formation of antibody or cell-mediated immunity against most 'self' components, analogous to the inherited non-reactivity of certain strains of animals to synthetic polypeptide antigens. Burnet has,

however, suggested an attractive alternative explanation, that the various potential antigens in an individual's tissues do act on the cells responsible for immune reactions but that, instead of causing antibody formation or cell-mediated immunity, they lead during fetal life to specific immunological tolerance. The subject is, however, a complex one; it is considered more fully on p. 132.

Route of administration of antigen. In most cases antigens elicit an immunological response only when they are introduced parenterally (i.e. not through the alimentary canal) so that their macromolecular state and the configuration of their antigenic determinants are not destroyed by digestion in the gut. Traces of certain proteins, such as those in heterologous milk, may, in fact, be absorbed from the gut and bring about specific sensitisation, especially in infants. In general, however, antigens introduced via the portal circulation are less immunogenic than when administered by other routes.

Antibodies

Antibody molecules have the special property of combining specifically with antigen or hapten. In so doing they may cover up harmful areas on molecules of toxin, in which case they are said to be **antitoxins**, or their combination with cells such as bacteria may lead (with the help of complement—p. 144) to death and lysis of the bacteria (**bacteriolytic effect**) or phagocytosis by polymorphs and macrophages (**opsonic effect**). Chemically, antibodies belong to the **immunoglobulin (Ig)** proteins of the plasma (for-

merly called γ -globulins because of their predominant electrophoretic mobility); and there are five classes: IgG, IgM, IgA, IgD and IgE. All immunoglobulins are composed of one or more similar units, each unit consisting of two pairs of identical polypeptide chains (Fig. 5.1); one pair, termed the **heavy chains**, are about twice the size (molecular weight) of the other pair, which are termed the **light chains**. Digestion of an Ig molecule by papain breaks it into three fragments, of which two are identical and

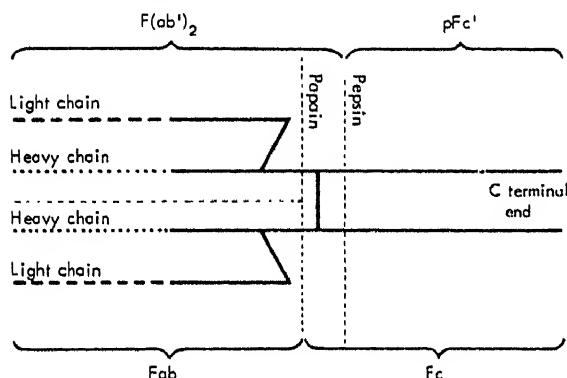


Fig. 5.1 Structure of monomeric immunoglobulin molecule. In any one molecule the two light chains have an identical amino-acid sequence and so also have the heavy chains. Each immunoglobulin class has a distinctive Fc piece (C terminal end of heavy chains). The amino-acid sequence of the interrupted portions of the light chains and dotted portions of heavy chains (the N-terminal ends) vary greatly among immunoglobulin molecules even of the same class and constitute the specific antigen-binding (Fab) sites, of which there are two on each molecule. The constitution of these variable regions of an Ig molecule is also known as the idotype. (The light lines indicate separation of the molecules into an Fc and two Fab fragments by papain digestion, and into one F(ab')₂ and two pFc' fragments by pepsin.)

are termed **Fab** (antigen-binding fragments) because each contains a combining site for antigen. The third fragment consists of the C-terminal ends of the heavy chains, and is termed **Fc** fragment; it is readily obtained in crystalline (hence Fc) form. As shown in Fig. 5.1, digestion of IgG by pepsin frees two pFc' fragments but leaves the two Fab fragments united by part of the Fc fragment as a single fragment—F(ab')₂. Heavy chains differ structurally for each class of Ig, and the letters γ , μ , α , δ , ϵ , are used to indicate the heavy chains of IgG, IgM, IgA, etc. respectively. By contrast, there are only two types of light chain, κ and λ , in all Ig classes, and each Ig molecule has *either* κ *or* λ light chains. Differences in the behaviour of antibodies of the same specificity, but of different Ig classes, and of the four sub-classes of IgG (IgG 1–4), are determined by the properties of the Fc part of their heavy chains.

The combination of antibody with antigen is believed to be achieved by hydrogen bonding, electrostatic, hydrophobic and van der Waal's forces; all of these are effective over a very short range and hold separate molecules together only when they fit snugly. The specificity

of antigen-antibody union therefore depends on the antibody combining-site having a complementary shape to the antigenic determinant which permits the necessary close fit. The shape of the combining site is determined by the amino-acid sequence of the N-terminal ends of the heavy and, to a lesser extent, of the light chains. In view of what is now known of protein synthesis this is of great interest and of fundamental importance in elucidating how antigenic stimulation gives rise to specific antibody formation. In contrast to most other proteins, each one of which in a given individual is of uniform amino-acid sequence, the immunoglobulins in a serum show marked heterogeneity of their N-terminal regions (the **variable regions**), the number of variations amounting to millions. This variety is responsible for the great range of antibodies which can develop in response to stimulation by an enormous number of different antigens.

Some idea of the specificity of antibody-antigen union can be gained from study of antibody against chemically defined haptens. Landsteiner, for example, showed that antibody raised against para-aminobenzene sulphonic acid does not combine with the ortho-form but gives a weak reaction with meta-aminobenzene sulphonic acid (Fig. 5.2). The

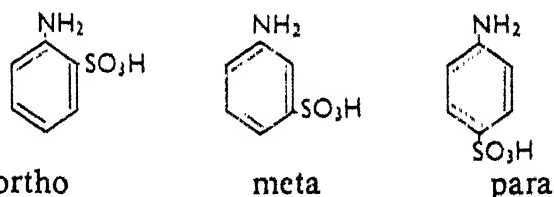


Fig. 5.2 Isomeric forms of aminobenzene sulphonic acid.

latter is called a cross reaction and it implies immunological reactivity with an antigen different from that which has led to the production of antibody. It results from the production of some antibody molecules which fit the cross reacting antigen sufficiently well to permit intermolecular attraction by short-range forces (see above). The closeness of fit between the antigen-binding sites of antibody with an antigen of given configuration can thus vary and this affects the firmness of combination; we therefore speak of high or low **affinity** of antibody for a given epitope. Cross reactions are generally of low affinity.

When an antiserum contains a variety of

antibodies reacting with multiple and often heterogeneous epitopes on a macromolecular antigen, the strength of the binding together of antigenic molecules by antibodies is referred to as the **avidity** of the antiserum; this is influenced not only by the affinities of each of the various antibodies reacting with individual epitopes, but also by the number of antigen-antibody linkages formed: although individual antibody-epitope linkages dissociate, those which fit snugly dissociate less readily than poorly-fitting linkages, and large numbers of firm linkages will result in the maintenance of avid binding.

Quite apart from their specific reactivity with antigens, antibodies of different classes possess properties which depend on the Fc part of the heavy (μ , γ , α , etc.) chains. These class distinctions are described on pp. 108-9.

The production of antibody

Injection of an antigen to which an individual has not previously been exposed results in a **primary antibody response**, i.e. the transient appearance in the blood of a small amount of specific antibody, mainly of IgM class, about seven days after the injection. Re-injection of the same antigen at a later date leads to a **secondary** or **anamnestic** (remembering) response in which large amounts of specific antibody, most of which is usually of IgG class, appear in the blood rapidly (in four days or so) and continue to be produced, although in gradually diminishing amounts, for weeks, months or even years. The greatly enhanced antibody production of the secondary response is the reason for the repeated injections of microbial antigens (vaccines) widely used in prophylactic immunisation.

Most of the antibody found in serum is produced by plasma cells in lymph nodes, spleen and bone marrow but some may also be formed by plasma cells in the lymphoid tissue of the gut and in the inflammatory lesion which forms around injected antigenic material. **Plasma cells** (Fig. 5.3) are ovoid cells, somewhat larger than small lymphocytes, and with a small round nucleus in which granules of chromatin are regularly spaced around the periphery, giving a 'cart-wheel' or 'clock-face' appearance. Their cytoplasm is basophilic and also pyroninophilic, indicating a high content



Fig. 5.3 Plasma cells in a lymph node draining a focus of infection. Note the eccentrically-placed, round nucleus with clumping of chromatin, and the deeply-stained (basophilic) cytoplasm showing, in some instances, a crescentic area of pallor alongside the nucleus. $\times 1000$.

of ribonucleic acid, and electron microscopy (Fig. 5.4) reveals a large amount of complex rough endoplasmic reticulum of the type found in cells which synthesise and secrete protein. It has been shown by immunofluorescence that each plasma cell at any given time produces light chains together with heavy chains of only one immunoglobulin class (e.g. IgG or IgM). Furthermore, *following stimulation with two distinct antigens (e.g. diphtheria and tetanus toxins) individual plasma cells will produce antibody to one or the other but not to both of these antigens.*

An additional polypeptide, the **J chain**, is synthesised by plasma cells producing IgM and IgA. It links together the basic 4-chain units of IgM and IgA to form polymers (see below).

Properties of the immunoglobulin classes

The various immunoglobulin classes have different functions (beyond that of specific com-



Fig. 5.4 Electron micrograph of part of the cytoplasm of a plasma cell, showing the abundant rough endoplasmic reticulum. Part of a mitochondrion is also included. $\times 108\,000$.

bination with antigen, which is common to all) and this is determined by the structure of the part of the heavy chains included in the Fc fragment.

When immunoglobulin of a particular class is injected into animals of another species it acts as an *antigen*, and antibodies specific for the light and for the heavy chains appear in the serum of the injected animal. Combination of the latter antibody with immunoglobulin *in vitro* provides a simple method of demonstrating the class to which a particular immunoglobulin belongs, for example by immunoelectrophoresis (p. 111).

Table 5.2 compares some of the features of the five known immunoglobulin classes. IgG is present in the plasma and extravascular spaces in the largest amount. IgG antitoxins are of importance because they combine with and neutralise toxins, thus protecting the individual from their harmful effects. There are receptors for the Fc part of IgG antibodies on polymorphs and macrophages which facilitate adherence and phagocytosis of the corresponding antigens. The reaction of IgG antibody with the

corresponding antigen can usually be demonstrated *in vitro* (see below); it can cross the human placenta and in this way passive immunity is transferred from mother to child. The four sub-classes of IgG (1–4) differ somewhat in their properties, e.g. capacity to acti-

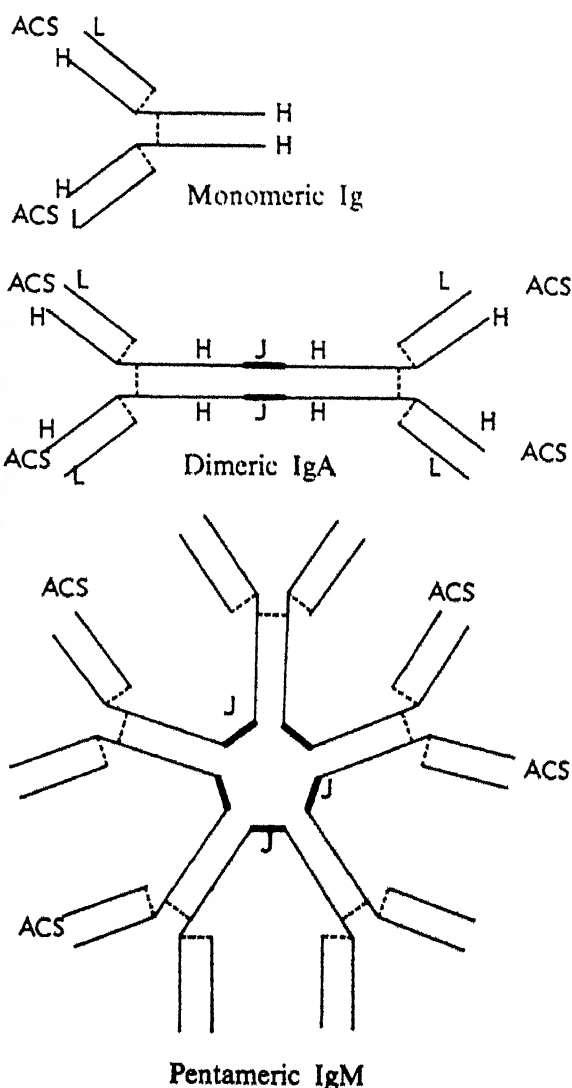


Fig. 5.5 The structure of immunoglobulins. The monomeric Ig molecule consists of two heavy and two light chains (*upper and Fig. 5.1*). IgA (*middle*) is secreted as a dimer, consisting of two monomeric units joined by J chains (heavy lines). Secretory IgA also contains a secretory piece (not shown). *Lower*, the pentameric IgM molecule, consisting of five monomers joined by J chains. Note that each unit has two antigen-combining sites (ACS), so that IgA dimer has four and IgM ten. (The heavy chains of Ig are angulated at the so-called hinge region, so that the unit molecule is Y-shaped: this is not shown in Fig. 5.1.)

vate complement (see below). These differences are referred to later in relation to topics in which they are of importance. **IgM**, a macroglobulin consisting of pentamers of the basic four-chain units or monomers (each of which has two antigen-combining sites) linked by J chains (Fig 5.5), is the first Ig class of antibody to be produced following the initial introduction of an antigen. Because of its 10 combining sites, it is usually of high avidity: it is especially effective in activating complement (see below), and its reactions are readily demonstrated *in vitro*. Except in inflammatory lesions, it is largely confined to the plasma, where it has an important function in destroying micro-organisms. The presence of IgM antibody, e.g. to rubella virus, indicates recent or continuing exposure to that antigen. **IgA** is secreted locally by plasma cells in the intestinal mucosa, the lacrimal glands, respiratory passages and salivary glands. It is secreted by the plasma cells in the form of a dimer, the two molecules being linked by a J chain. Such locally produced IgA is taken up by the glandular and lining epithelial cells of these tissues and coupled with a carbohydrate 'transport piece' which renders it more resistant to digestive enzymes. In this form it is secreted in the tears, saliva, alimentary and salivary mucus, etc. and forms a lining, sometimes referred to

as 'antiseptic paint', over the alimentary and respiratory mucous membranes and conjunctiva. IgA activates complement by the alternative pathway and may enhance the bacteriolytic activity of lysozyme (p. 175). It is also present in the plasma in the form of monomers and dimers. **IgE** has the special property of attaching to tissue, particularly to mast cells and basophils, by means of its Fc fragment, leaving the specific combining sites (on the Fab fragments) available for union with antigen. If such union takes place, pharmacologically active substances such as histamine are released, with the production of an anaphylactic hypersensitivity reaction within a few minutes (p. 146). The biological properties of **IgD** are unknown, but it acts as a lymphocyte surface antigen-receptor (pp. 125, 128).

Table 5.2 Size and plasma (or serum) concentrations of immunoglobulins

Class	Molecular weight (daltons)	Degree of polymerisation	Concentration (normal serum) g/litre
IgG	150 000	Monomer	8-16
IgM	900 000	Pentamer	0.5-2
IgA	Mainly 150 000	Mono- and dimer	1.4-4
IgD	185 000	Monomer	0-0.4
IgE	200 000	Monomer	$2-45 \times 10^{-7}$

Demonstration of antigen-antibody reactions

This section and the following one on demonstration of cell-mediated immunity are largely of a technical nature and are intentionally brief. To some extent, they interrupt the account of the immune response which continues on p. 114. They do, however, illustrate the practical aspects of what might otherwise seem to be largely an academic subject. Accordingly, the reader is advised not to skip them.

The demonstration of antigen-antibody reactions is applicable equally to the detection and assay of antigen by means of a known antibody and of antibody by means of a known antigen.

Antigen-antibody reactions in vivo may be demonstrated in three ways.

(1) A potentially harmful antigenic substance

may be rendered harmless by union with antibody and not have the expected effect. For example, in the Schick test, intradermal injection of a small amount of diphtheria toxin into the skin of a non-immune individual results in an area of inflammation. The diphtheria antitoxin present in an immunised individual neutralises the toxin and so suppresses the inflammation. Similarly, antibodies to pathogenic micro-organisms may be demonstrated by their capacity to protect experimental animals against a lethal dose of the micro-organism.

(2) A normally harmless stimulus may result in tissue injury, i.e. a hypersensitivity reaction. For example, inhalation of grass pollen may induce an attack of hay fever or asthma, mediated by its reaction with antibody specific

for grass pollen. In this instance, the antibody is usually of IgE class, but hypersensitivity reactions may result from the union *in vivo* of other classes of antibody with antigen: they are of considerable importance in disease processes and are described in Chapter 6.

(3) In the special instance of antibodies which react with antigenic constituents of the host cells, death of the target cell may result, e.g. destruction of lymphocytes by anti-lymphocyte serum. There are also rare examples of antibodies which alter the physiological activity of target cells, e.g. increased secretion of thyroxine as a result of the reaction of auto-antibody specific for the TSH-receptor of thyroid epithelium. By binding to the receptor, the antibody has the same effect as TSH.

Antigen-antibody reactions *in vitro* may be demonstrated in various ways, depending on the nature of the antigen and the type and amount of antibody present.

(1) **Visible aggregation of antigen.** Since each antibody molecule has at least two combining sites it can bind with two or more antigen molecules. If the antigen molecules contain several antigenic determinants, and if they are in solution, antibody can form cross-linkages between antigen molecules, uniting them in the form of a lattice (Fig. 5.6): if antigen and antibody molecules are present in optimal combining proportions the aggregates will be large (Fig. 5.6a) and visible as an insoluble precipitate (**precipitin reaction**). Lattice formation and therefore precipitation can be inhibited when an excessive amount of antigen saturates the combining sites on all the antibody molecules, and smaller complexes may then be formed (Fig. 5.6b). Conversely, when gross excess of antibody is present (Fig. 5.6c), each antigen combining site may fix a separate antibody molecule, and, once again, the complexes may be too small to form a visible precipitate (**prozone effect**). Optimum antigen-antibody proportions for lattice formation can readily be achieved by allowing antibody and antigen to diffuse towards each other through agar (Ouchterlony technique, Fig. 5.7). The interpretation of agar diffusion tests involving antiserum raised against a complex mixture of antigens (as when human serum is injected into rabbits) is facilitated by partial separation of the constituent antigens by electrophoresis prior to the precipitin reaction; this method, called

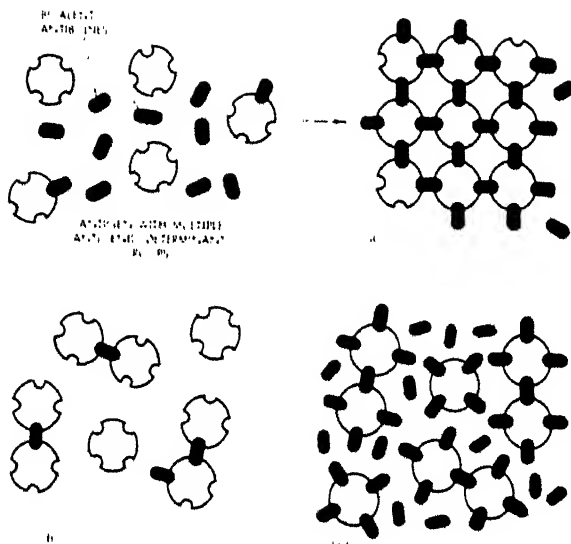


Fig. 5.6 The reaction of antigen and antibody (a) in optimal proportions to form large aggregates: (b) in antigen excess: (c) in antibody excess. In (b) and (c) small complexes are formed. In the case of soluble antigens, large aggregates form as in (a), and produce a precipitate, whereas with particulate antigens, e.g. bacteria, formation of aggregates is termed agglutination.

immunoelectrophoresis, is illustrated in Fig. 5.8.

When the antigen molecules are associated with a large particle, e.g. the antigenic determinants of the surface of a red cell or bacter-

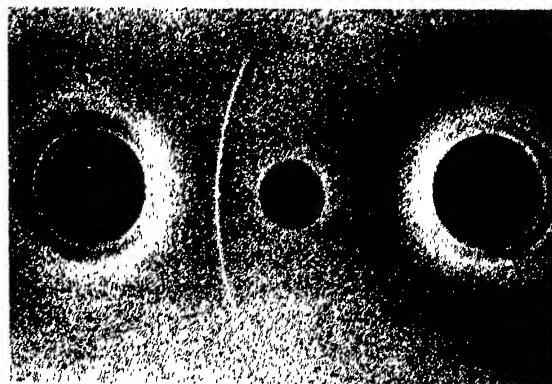


Fig. 5.7 Ouchterlony technique. The central well contains a solution of antigen. The well on the left contains the corresponding antiserum, and that on the right a negative control serum. A white line of precipitate, composed of antigen-antibody complex, has formed between the antigen and antibody wells. In this instance the antigen is thyroglobulin and the test detects auto-antibody to thyroglobulin in the serum of a patient with chronic thyroiditis.

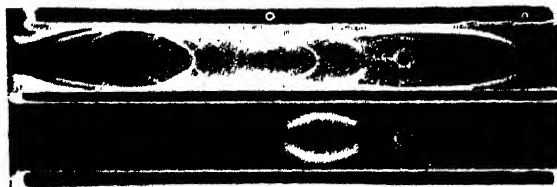


Fig. 5.8 An illustration of the use of immunoelectrophoresis to detect antigens. A mixture of antigens (in this case human serum in the upper well) is subjected to electrophoresis in agar. Antiserum (rabbit antiserum to whole human serum) is then placed in the trough and diffusion allowed to proceed. Each antigen reacts with the corresponding antibody to form an arc of precipitation, the position of which depends on the electrophoretic mobility of the antigen. Use of a single purified antigen (in this case the C3 component of human complement in the lower well) helps to determine whether that antigen is present in the mixture. The upper and middle troughs contain antiserum to whole human serum and the lower trough antiserum to human C3.

ium, or are adsorbed artificially onto red cells or latex particles, antibody causes aggregation of the particles (**agglutination reaction**). The visible aggregation of large particles such as bacteria can be effected by minute amounts of antibody.

(2) **Methods using anti-immunoglobulin.** In certain circumstances antibody combines with antigen without causing aggregation. This may occur if the spatial arrangement of the antigen (e.g. the Rhesus antigen on the surface of red cells) prevents the divalent IgG antibody molecule from combining simultaneously with antigenic determinant groups on two different red cells. In these circumstances, exposure of the Rhesus positive red cells to anti-Rhesus antibody merely results in their being coated with IgG. However, the coated cells can be agglutinated by antibody against IgG (**antiglobulin or Coombs' reaction**) (Fig. 5.9).

A similar principle is used in the **indirect immunofluorescence (indirect fluorescent antibody) technique** to detect insoluble antigen (e.g. bacterial capsular polysaccharide) in a histological section or smear. When this is exposed to antiserum the antibody combines with antigen without visible effect. The slide is washed to remove the antiserum, leaving only the specific antibody which is, of course, an immunoglobulin, attached to the antigen present in the section. Antibody to immunoglobulin, conjugated with a fluorescent dye such as

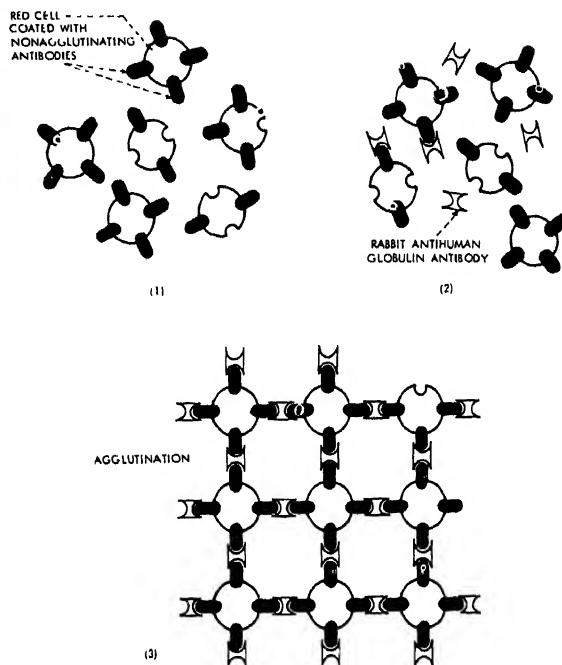


Fig. 5.9 Antigenic particles sensitised with a non-agglutinating antibody are agglutinated by antibody to immunoglobulin (antiglobulin reagent).

fluorescein isothiocyanate, is then applied to the section or smear and the site of antigen-antibody combination is visualised by the presence of the fluorescent antiglobulin when the section is examined microscopically with ultra-violet light (Figs. 5.10, 5.11). In the **direct immunofluorescence technique** the tissue or cells are washed to remove free immunoglobulins and treated directly with fluorescein-conjugated antiglobulin reagent. This demonstrates the binding of antibody (or the deposition of antigen-antibody complexes) *in vivo*. More recently, analogous techniques have been introduced in which antibody is labelled with the enzyme *peroxidase* instead of with a fluorescent dye; peroxidase can be localised as a dark deposit by ordinary light microscopy after appropriate histochemical treatment.

(3) **Methods using radioactive antigen.** In the Farr technique, which is valuable in measuring antibody to soluble antigen, excess antigen labelled with radioisotope is added to antiserum; the immunoglobulin (including antigen-antibody complexes) is then precipitated by 50 per cent saturation with ammonium sulphate. Provided that the free antigen (uncombined with antibody) is not salted out by this

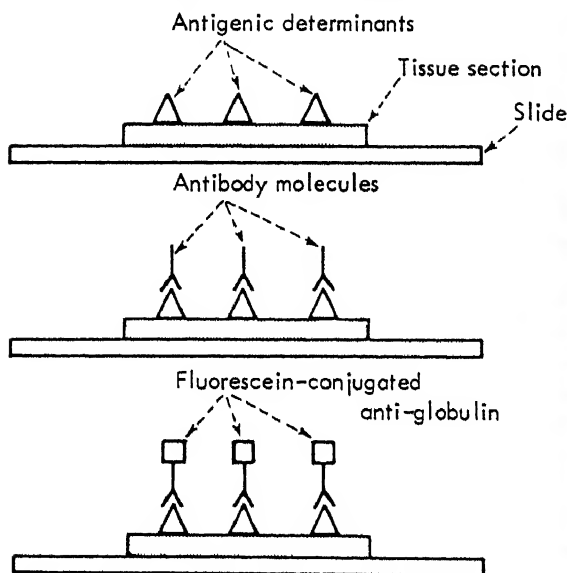


Fig. 5.10 The indirect immunofluorescence technique performed on a tissue section. *Top*, tissue section on slide. *Middle*, section treated with antibody (λ) to a tissue constituent and washed; antibody molecules adhere to the tissue antigen. *Bottom*, section treated with fluorescein-conjugated antibody (\square) to immunoglobulin: sites of antigen-antibody reaction fluoresce in ultraviolet light.

procedure, the amount of radioactivity precipitated is proportional to the amount of antibody in the serum.

In the **antiglobulin coprecipitation technique** the amount of antibody to a given radioactive antigen can similarly be measured by precipitation with class-specific anti-Ig antibody instead of by a salting out procedure.

Numerous very sensitive radioimmunoassays have been devised for the measurement of antigen or antibody. For example, the concentration of insulin in the plasma may be measured by mixing the plasma with anti-insulin and radioactive insulin in standard amounts, and determining how much radioactive insulin is excluded, by the plasma insulin, from combining with antibody.

(4) Methods involving damage to cells. Antibody combined with antigen on the surface of intact cells such as bacteria or erythrocytes can damage the cell membrane and cause lysis (partial dissolution) of the cells which is readily demonstrable. The lysis is mediated by a complex group of at least twenty factors present in fresh normal serum and known collectively as **complement** (p. 142). These factors are activated by the Fc portions of IgM and IgG anti-

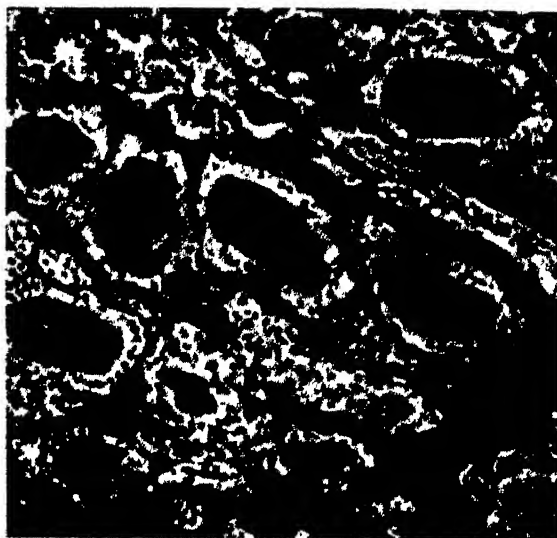


Fig. 5.11 A positive indirect immunofluorescence test for antibody to thyroid epithelium. A frozen section of thyroid tissue has been treated with the serum undergoing test (from a patient with chronic thyroiditis), followed by treatment with fluorescein-conjugated anti-human-IgG. Note fluorescence of the thyroid epithelial cytoplasm. (U.V. microscopy.)

bodies which have combined with antigen, and once activated they give rise to a chain of enzyme reactions culminating in digestion of the cell membrane where the antibody is attached. In the course of the reaction complement is used up, or 'fixed'.

(5) Complement fixation. Complement is fixed in many antigen-antibody reactions in addition to those involving cell-surface antigens. Invisible antigen-antibody reactions can often be demonstrated indirectly by allowing them to occur in the presence of a measured amount of complement and subsequently adding an indicator system, namely red cells coated with a 'haemolytic' antibody i.e. one which causes red cell lysis only in the presence of complement. Fixation of complement by the invisible antigen-antibody reaction prevents haemolysis in the indicator system (Fig. 5.12). A more detailed account of complement is given on p. 142.

(6) Blocking antibody. In certain circumstances (usually poorly defined), antibody can combine invisibly with antigen which thereby becomes unavailable for participation in a subsequent visible immunological reaction such as agglutination or cytotoxic damage. The presence of blocking antibody in a serum can thus be demonstrated by a two-stage test.

Mix antigen and antibody		Add complement (C)	Add sensitised RBC
Positive test	$Ag^x + Ab^x \longrightarrow Ag^x - Ab^x$ (union)	$+ C \longrightarrow Ag^x - Ab^x$ C (complement used up)	No lysis
Negative test	$Ag^x + Ab^y \longrightarrow Ag^x + Ab^y$ (no union)	$+ C \longrightarrow Ag^x + Ab^y + C$ (complement not used)	Lysis

Fig. 5.12 The complement fixation reaction depends on the 'fixation' of complement by an antigen-antibody complex (upper line). The fixation of complement is demonstrated by non-lysis of subsequently added red cells coated with a haemolytic antibody. If there is no antigen-antibody reaction (lower line), complement is not used up and lyses the sensitised red cells.

Demonstration of cell-mediated immunity

As noted on p. 103, immune responses to most antigens induce not only the production of antibody but also cell-mediated immunity (CMI), i.e. the production of specifically primed lymphocytes which are capable of reacting with the inducing antigen. When such lymphocytes encounter the antigen, they release biologically-active compounds termed *lymphokines* (p. 157), one of which causes an acute inflammatory reaction, while another immobilises macrophages. The reacting lymphocytes also enlarge and undergo mitosis and stimulate mitosis in other lymphocytes. Tests for CMI are based on such changes. It is demonstrated most readily *in vivo* by intradermal injection of the antigen, which leads to an indurated erythematous lesion in the dermis maximal in 24–72 hours—the **delayed hypersensitivity reaction**. During the development of the lesion there is early but transient emigration of polymorphs, followed by an increasing accumulation of lymphocytes and some macrophages around venules, hair follicles and sweat glands: this is accompanied by inflammatory hyperaemia and oedema (Fig. 6.8, p. 158). CMI may also be demonstrated in animals by injecting the antigen into the cornea which becomes opaque 24–48 hours later as a result of migration of mononuclear cells into it. These sites are convenient for eliciting delayed hypersensitivity reactions, but all tissues are reactive, for the sensitised small lymphocytes circulate through the blood and lymphoid tissues and can migrate into any tissue.

A classical example of CMI is that which develops to tuberculin in most individuals who are (or have been) infected with *Mycobacterium tuberculosis* or following inoculation with BCG (an attenuated strain of *Mycobacterium bovis*). The Mantoux test (p. 157), in which more or less purified tuberculin protein is injected intradermally, is a classical delayed hypersensitivity reaction.

Skin tests with appropriate antigens may be used similarly to test for CMI to various other micro-organisms, including viruses and fungi. They are also useful to elicit CMI to certain simple chemicals which, when applied to the skin, act as haptens and by combining with skin proteins become immunogenic: examples include *p*-phenylene diamine (in some hair dyes) and nickel salts (derived from nickel clasps on underwear, etc.): once CMI has developed, further application of the chemical to the skin results in an inflammatory lesion known as *contact dermatitis*, which is simply a delayed hypersensitivity reaction.

CMI may also be demonstrated *in vitro* by observing the effect of the appropriate antigen on living cells from an immune individual. For example, if leukocytes are placed in a capillary tube immersed horizontally in tissue culture medium, the cells migrate from the open end of the tube: such migration is inhibited by adding the antigen to the tissue culture medium. Secondly, when lymphocytes from the blood of an immunised individual are maintained in culture in the presence of the antigen, a pro-

portion of them transform into larger cells (lymphoblasts) with basophilic cytoplasm, synthesise DNA and undergo mitosis. Addition of radio-active thymidine to the culture medium and subsequent determination of its incorporation into DNA is a method of assessing the transformation. Such *in vitro* methods are being used increasingly to demonstrate CMI to various micro-organisms and other antigens.

A further method of determining the capacity of experimental animals to develop CMI consists in the application of a tissue allograft or xenograft; CMI normally develops to such foreign tissue and plays an important role in its destruction by the host: failure to reject a foreign skin graft is thus evidence of deficient CMI.

In experimental work it is sometimes useful to induce in animals CMI to soluble proteins and other antigens which, when injected alone, usually induce antibody production but only feeble CMI: the latter may be enhanced greatly

by mixing the antigen with killed *Mycobacterium tuberculosis*, other mycobacteria, or a peptidoglycolipid extracted from the walls of such organisms and injecting it in the form of an emulsion containing droplets of light mineral oil (Freund's type of 'adjuvant').

It should be noted that the demonstration of CMI *in vivo* is by eliciting a delayed hypersensitivity reaction which is a destructive lesion and may, if severe, progress to necrosis. Delayed hypersensitivity reactions occur naturally, e.g. in various infections and in contact dermatitis, and are responsible for most of the tissue injury in tuberculosis. CMI does, however, afford protection against various micro-organisms, notably viruses, fungi and many of the bacteria which cause chronic infections. This is well illustrated by children with congenital agammaglobulinaemia who cannot produce antibodies and yet overcome most viral infections normally unless they are also deficient in CMI.

The Cellular Basis of the Immune Response

So far, this account has concentrated mainly on describing the usual products of antigenic stimulation, i.e. antibodies and specifically-primed lymphocytes, capable of reacting with the antigen. We must now consider in detail the cellular events involved in these responses.

The cytology of immunological phenomena is complex but is worth studying in some detail, not only because of the vital importance of specific immunity, but also because it illustrates the complexity of the systems by which eukaryotic cells communicate with one another and co-ordinate their activities. Most of the advances have been based on the manipulation of cells and tissues of experimental animals, but the features of immune responses and of naturally occurring immunodeficiency states in man, the effects of therapeutic immunosuppression, and investigation of human lymphocytes *in vitro* all indicate that the same basic rules apply.

The main features of the cellular basis of antibody production and cell-mediated immunity may be summarised as follows.

1. Cell-mediated immunity and antibody production are both attributable to lymphocytes capable of recognising and responding specifically to the stimulus provided by an antigen, i.e. **specifically-responsive lymphocytes**.
2. At some stage in their development, lymphocytes become '**committed**', i.e. capable of responding only to a particular antigenic determinant group, or to closely similar determinant groups. Lymphocyte populations thus consist of individual cells which differ in the antigens to which they can respond. In consequence, no one antigen can stimulate a response in more than a small proportion of them.
3. The specifically responsive lymphocytes which bring about cell-mediated immunity are **thymus-dependent** or **T lymphocytes**. They develop from stem cells in the thymus under the influence of a thymic hormone and many of them leave the thymus and reach the various other lymphoid tissues.

4. On encountering an antigen to which it is specifically responsive, the T lymphocyte proliferates in the lymphoid tissues to produce a clone of lymphocytes, all of which are capable of reacting with that antigen—**specifically-primed T lymphocytes**. These are responsible for delayed hypersensitivity reactions.
5. The specifically-responsive lymphocytes which bring about antibody production are **thymus-independent** and are termed **B lymphocytes**. In mammals, they develop from stem cells in the haemopoietic tissue from which many of them migrate and pass to the various other lymphoid tissues.
6. On encountering an antigen to which it is specifically responsive, the B lymphocyte, like the T lymphocyte, proliferates in the lymphoid tissues to form a clone of cells capable of reacting with that antigen. Some cells of the clone differentiate into plasma cells and secrete antibody which also is capable of reacting with the antigen. The plasma cells are thus derived from B lymphocytes.
7. Some of the lymphocytes produced by antigen-induced proliferation of T and B lymphocytes persist as **memory cells**. In other words, antigenic stimulation increases the number of T and B lymphocytes capable of reacting with that antigen, and some of these lymphocytes are long-lived and are responsible for a secondary response on subsequent stimulation by the antigen.
8. Although T lymphocytes do not give rise to antibody-producing plasma cells, they co-operate with B cells in antibody responses, and such co-operation enhances the production of antibody to most antigens, and is essential for some antibody responses. T lymphocytes can also have a suppressive effect on immune responses.
9. In some circumstances, antigenic stimulation results in neither antibody production nor cell-mediated immunity, but renders the individual specifically unresponsive to subsequent challenge by that antigen. This unresponsive state is termed **acquired immunological tolerance**. It provides an explanation of why we do not usually respond strongly to antigens in our own cells and tissues, and it plays an important

role in successful transplantation of foreign cells and tissues.

These basic features of immune responses are considered more fully below.

The specifically responsive lymphocyte

For the development of specific immunity, it is necessary that cells should recognise specific determinant sites (epitopes) on antigenic molecules, and should respond to them in ways that lead to the production of antibodies and primed lymphocytes, both of which are capable of reacting specifically with the antigen. Much of the credit for demonstrating that these cells—the keystones of the immune response—are lymphocytes, is due to Gowans and others (see Gowans, 1966). They used techniques in which lymphocytes were removed from rats by thoracic duct drainage. By drainage for several days, the cell-free fluid being returned to the rat, depletion of a population of small lymphocytes was achieved and the depleted rats were found to be defective in their responses to antigenic stimulation. For example, when challenged by injection of sheep erythrocytes or tetanus toxoid they made poor antibody responses, and when grafted with skin from an allogeneic rat they did not reject the graft, indicating failure of the normal cell-mediated response to the graft antigens. These immunological deficiencies were corrected by injection of small lymphocytes from the thoracic duct of a normal rat, and when the donor had been previously immunised, e.g. by a skin allograft or an injection of sheep erythrocytes, the recipient showed the rapid and intense response to antigenic challenge which is characteristic of the secondary response. These observations suggested that small lymphocytes in thoracic duct lymph are essential for primary immune responses, and showed more conclusively that they include specifically responsive cells resulting from a previous immune response, i.e. immunological memory cells (thus explaining the rapidity of the secondary response). Gowans' findings have since been confirmed by many other workers and have been shown to apply to several mammalian and avian species. Animals depleted of lymphocytes by various other methods (p. 121) have also been shown to be immunologically deficient and the deficiency is corrected by injection of lymphocytes.

tion of lymphocytes from normal animals. Moreover, in such experiments it has been shown by cell labelling techniques that the donated lymphocytes proliferate and give rise to the plasma cells which produce antibody and the lymphocytes responsible for cell-mediated immunity.

By such experiments, it has been firmly established that, in both primary and secondary immune responses, the cells capable of recognising the antigen and of responding specifically to it by production of antibody or cell-mediated immunity, are lymphocytes.

The origin of lymphocytes: the primary lymphoid organs

Evidence is presented below that the thymus is the original source of a major population of lymphocytes termed **thymus-dependent** or **T lymphocytes**, and that there is a second population of lymphocytes which, in birds, arises in a pouch-like outgrowth of the hind-gut, the *bursa of Fabricius*. In mammals, this second population, termed **B lymphocytes**, originates mainly, and probably solely, in the haemopoietic tissue. Production of lymphocytes in the thymus and haemopoietic tissue occurs spontaneously, being quite independent of antigenic stimulation, and they are commonly called the **primary lymphoid organs**.

The origin of T lymphocytes: the thymus

Development. The thymus develops from bilateral epithelial ingrowths of the endoderm of the 3rd and 4th branchial arches accompanied by mesenchyme which is probably of ectodermal origin. In most mammals, these ingrowths fuse to form a single organ. Meanwhile some of the haemopoietic stem cells, developing in the yolk sac and migrating into the blood, settle in the thymus where they proliferate rapidly and differentiate into lymphocytes. The thymus now appears as a lobulated organ (Fig. 5.13) in which the outer or cortical layer is crowded with proliferating lymphocytes lying in a network of epithelial cells. In the inner or medullary layer, lymphopoiesis is less active and there are aggregates of epithelial cells (Hassall's corpuscles).

Colonisation of the thymus by stem cells from the blood has been demonstrated by tissue-culture studies of embryonic thymic tissue and by thymus grafting experiments. For example, Le Douarin (1977) and others showed that when the epithelial thymus

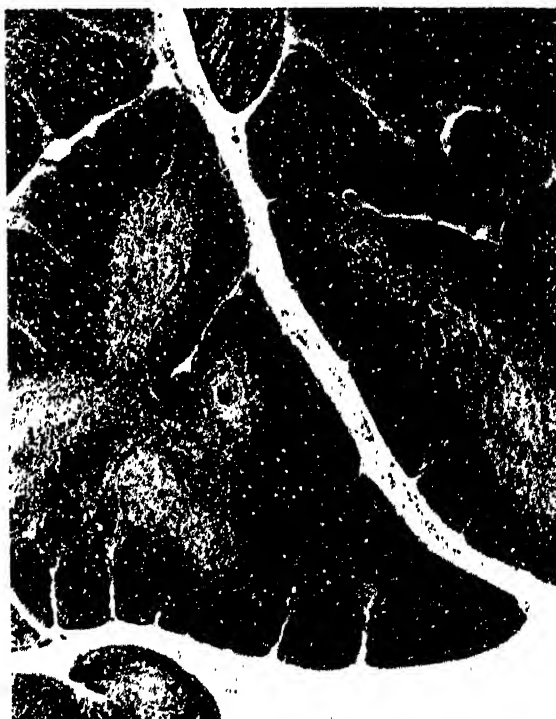


Fig. 5.13 Thymus in childhood, showing lobulation and division into cortex (darker areas) and medulla. $\times 30$.

of an early Japanese quail embryo was implanted for two days in a chick embryo,* and then transferred back to a quail embryo, the lymphocytes first developing in it were of chick origin, derived from stem cells in the blood which settled in the graft during its sojourn in the chick embryo. Subsequently stem cells from the second (quail) host settled in the graft, and quail lymphocytes developed in it.

By other cell transfer experiments it has been shown that stem cells capable of colonising the thymus originate in haemopoietic tissue, appearing first in the yolk sac, then in the fetal liver and subsequently in the bone marrow.

Similar experiments with mice, using chromosomal morphological markers to distinguish between

* These two species were chosen because their lymphocytes can be distinguished morphologically.

donor and host lymphocytes, have led to the same conclusions.

Function. Until quite recently the thymus was an organ of mystery. Unlike the other lymphoid tissues (lymph nodes, spleen, tonsils, etc.) it is not a site of significant immune responses following exposure to antigen, and the proliferation of lymphocytes in the thymic cortex, which is maximal around the time of birth but continues throughout life, is quite independent of antigenic challenge. This suggested that the thymus supplies lymphocytes, but many lymphocytes produced in the cortex die *in situ*, their nuclear debris being conspicuous in macrophages, and export of thymic lymphocytes remained for long no more than a likely possibility. Nor did early studies of thymectomy elucidate thymic function, for no important immunological effects were observed.

The important clue to thymic function came from the observation by Good and his co-workers (1964) and others that congenital immunological deficiency states in man were sometimes associated with thymic abnormalities. Miller (1964) then demonstrated that mice thymectomised *at birth* failed to thrive and commonly died of a wasting disease since shown to be due mainly to infections. The thymectomised mice were shown to be defective in cell-mediated immune responses and in their antibody responses to some antigens. They were also deficient in lymphocytes in the blood, thoracic duct lymph, and parts of the lymph nodes, spleen, tonsils, etc. These deficiencies were soon shown to be prevented, or completely and permanently corrected, by a syngeneic thymus graft, while lymph node or spleen cells from normal mice were partially restorative. Similar deficiencies occur spontaneously in an inbred strain (*nu nu*) of mice with thymic aplasia.

It thus appears that, in mice, the presence of the thymus is necessary for the development of a major population of lymphocytes which are responsible for cell-mediated immune responses and which also influence antibody production. By preparation and use of suitable antisera, mouse thymocytes have been shown to exhibit various surface antigens, some of which (e.g. θ or Thy-1 antigens) are shared by many of the lymphocytes in the blood, lymph and other lymphoid tissues. It is these lymphocytes with 'thymic markers' which fail to develop following neonatal thymectomy or in *nu nu* mice;

accordingly they are termed the **thymus-dependent** or **T lymphocytes**, and mice lacking them are sometimes termed '*B*' mice.

The same immunological and T-cell deficiencies as occur in mice have been shown to follow neonatal thymectomy in several other species, and can be prevented or corrected by thymus grafts. In man, the T lymphocyte population develops long before birth, but congenital failure of thymic development (p. 170) is associated with the same features as neonatal thymectomy in mice.

It has also been demonstrated in a number of species that, in conditions much closer to the physiological state than in the experiments described above, the thymus supplies lymphocytes to the other lymphoid tissues. This was done by injecting ^3H -thymidine into the thymus, so that it became incorporated in the nuclei of dividing thymocytes: labelled lymphocytes were subsequently detected by auto-radio-graphy in the other lymphoid tissues.

Thymic function in the adult. As the thymus slowly involutes following puberty, its lymphopoietic activity diminishes but does not cease entirely, even in old age. The diminishing importance of the thymus is illustrated by performing thymectomy in mice after the neonatal period: the longer it is delayed, the less its effect. This is clearly because the thymus provides much of the T-lymphocyte population shortly after birth. Even in adult mice, however, thymectomy has some effect, for a year or more later the mice show deficient cell-mediated immunity as illustrated by reduced capacity to reject a skin allograft. Similarly in man, thymectomy in young adults (performed therapeutically in patients with myasthenia gravis) has no immediate immunological effect, but defective cell-mediated immunity and a diminution in T lymphocytes in the blood has been reported in patients examined 15 or more years later.

In spite of its involuted state in the adult, the thymus remains capable of replacing the whole T-lymphocyte pool. This is demonstrated by subjecting mice to a dose of x-irradiation which destroys virtually all the body's lymphocytes and haemopoietic stem cells, and administering bone marrow cells to restore the haemopoietic tissue. Regeneration of the T lymphocytes (from haemopoietic stem cells) is dependent on the presence of the thymus and is prevented by concomitant thymectomy.

The thymus thus produces and exports T

lymphocytes (Fig. 5.14). This occurs maximally during fetal life or shortly after birth (depending on species). Thymectomy before production of the T-lymphocyte population results in a virtual absence of T cells and consequent severe immunological deficiencies. The thymus continues to supply T lymphocytes after birth but this function becomes progressively less important with age, the T-lymphocyte population, which includes much of the individual's immunological experience stored in memory cells, becoming increasingly self-sufficient.

Thymic lymphopoietic hormone. There is evidence that the thymus produces a hormone, 'thymosin', which stimulates differentiation and proliferation of T lymphocytes. This was first suspected when implantation of thymic tissue enclosed in cell-proof diffusion chambers was found to restore partially the T-lymphocyte population of neonatally thymectomised mice. Subsequently, extracts of thymic tissue have been shown to have a similar effect, and similar activity has been detected in normal human and mouse plasma, but is absent following thymectomy. According to Bach, *et al.* (1975), the plasma activity in man decreases with age, an

observation which accords well with the normal process of thymic involution. It thus seems likely that production of T lymphocytes in the thymus is stimulated by the local effect of thymic hormone.

Although administration of thymus extract has some influence on differentiation of T lymphocytes in the thymectomised animal, there is no strong evidence that production of mature T lymphocytes from stem cells can occur without the thymus.

From electron-microscopic and other observations, the thymic epithelium seems the most likely source of thymosin.

The origin of B lymphocytes

In birds, the *bursa of Fabricius*, like the thymus, develops as an epithelial organ and is invaded in embryonic life by haemopoietic stem cells which proliferate and differentiate into lymphocytes: these, in turn, enter the blood and help to populate the lymph nodes, spleen, etc. They constitute a second major population of lymphocytes, termed **B lymphocytes** (B for bursa-dependent), and bursectomy of the embryo

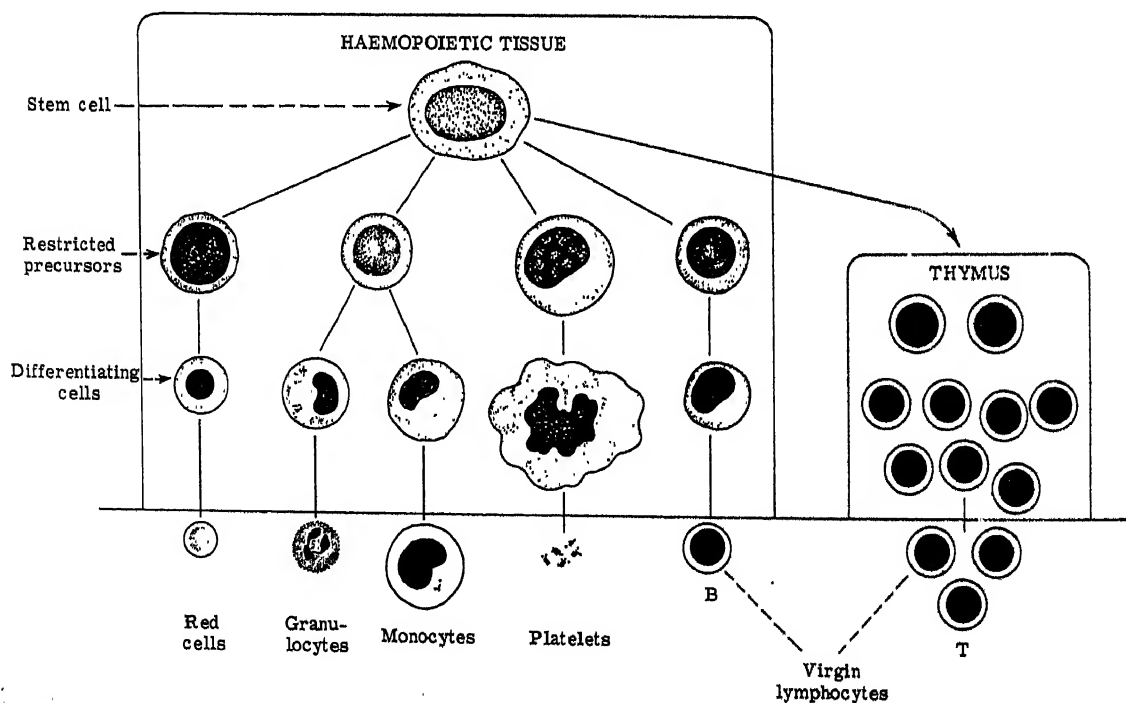


Fig. 5.14 In addition to differentiating into restricted precursors of red cells, granulocytes, monocytes and platelets, the haemopoietic stem cell differentiates into pre-B cells which undergo B-lymphopoiesis (in mammals within the haemopoietic tissue) and provide virgin B lymphocytes. Some stem cells pass into the blood and settle in the thymus, giving rise to virgin T lymphocytes.

results in failure to develop B lymphocytes. Such birds develop T cells normally, and can make cell-mediated immune responses, but they lack plasma cells and cannot make antibodies; accordingly, the plasma is devoid of immunoglobulins ('agammaglobulinaemia').

The haemopoietic marrow of mammals contains large numbers of lymphocytes which are

rapidly labelled by ^3H -thymidine (p. 117), i.e. they are an actively dividing population, and the labelled immature cells mature into B lymphocytes, some of which can subsequently be detected in the lymph nodes, spleen, etc. *Haemopoietic tissue is thus a site of production of B lymphocytes from stem cells* (Fig. 5.14), but it is not known whether it is the only site.

Lymphocytes in the secondary lymphoid tissues, blood and lymphatics

In contrast to the primary lymphoid organs, in which generation of lymphocytes is independent of antigenic stimulation, *lymphocyte proliferation in the secondary lymphoid tissues (lymph nodes, spleen, tonsils, Peyer's patches, etc.) is a response to antigenic stimulation.*

The secondary lymphoid tissues receive from the primary lymphoid tissues T and B lymphocytes which have not yet responded to antigenic stimulation ('virgin' lymphocytes). When antigen is introduced into the body, those virgin lymphocytes capable of responding to it undergo a clonal proliferation, giving rise to the effector and memory cells of the primary immune response. The **effector cells** are those cells responsible for the defensive mechanisms of specific immunity; they include the plasma cells which secrete antibody and the cytotoxic T lymphocytes which react directly with antigen and mediate the delayed hypersensitivity reaction. **Memory cells** of T and B origin are small lymphocytes specifically primed to react with the antigen which has induced their production. On second or subsequent introduction of a particular antigen into the body, the immune response is augmented by participation of T and B memory lymphocytes in the secondary lymphoid tissues: these cells give rise to large numbers of effector cells and also memory cells, and are responsible for the rapidity and intensity of secondary immune responses.

The secondary lymphoid tissues (lymph nodes, spleen, tonsils, Peyer's patches, etc.) contain a complex mixture of T and B cells, and the situation is further complicated by the continuous redistribution of lymphocytes via the blood and lymphatics. Considerable progress has, however, been made in defining and identifying the major lymphocyte sub-

populations in the secondary lymphoid tissues and the blood, and in elucidating their lifespan, function and migratory behaviour. The main features are described briefly in the following sections, but first it is necessary to note some of the methods of recognising T and B lymphocytes.

Identification of T and B lymphocytes

Although morphologically similar, T and B lymphocytes differ in a number of ways. It is, however, important to appreciate that the features of both types of cell change during their life cycle. The situation is further complicated by the existence of subsets of lymphocytes with features not characteristic of either T or B cells, the nature of which is uncertain. Some of the features used for identification are summarised below.

Surface antigens. Depending on their degree of maturity, the lymphocytes of various species, including man, may be identified by hetero-antisera specific for T and B cells. Such antisera may be used to identify individual cells by immunofluorescence or, together with complement, to destroy either T or B cells *in vivo* or *in vitro*.

Intracellular esterase activity is readily demonstrable in T, but not in B, lymphocytes.

Response to mitogens. A number of substances induce mitosis preferentially in T or B cells. For example, phytohaemagglutinin and concanavalin A induce mitosis mainly in T cells while bacterial lipopolysaccharides induce B-cell mitosis. Using these and other reagents, *in vitro* mitotic activity may be assessed by incorporation of ^3H -thymidine, but these tests are used mainly as indirect indications of T or B cell functional activity and are unsuited to identifying individual lymphocytes.

Surface immunoglobulin. As they mature in the haemopoietic marrow, B lymphocytes develop surface immunoglobulin; this is detectable in both

virgin and memory B cells, and also up to a late stage of differentiation of B cells into plasma cells. Surface Ig may be detected by use of anti-Ig antisera (Fig. 5.15).

Receptors for Fc. B cells and K cells (see below) have surface receptors for the Fc part of IgG which has been aggregated by heat or complexed to an antigen. Fc receptors may be demonstrated by the binding of antibody-sensitised red cells to form rosettes (Fig. 5.16) or fluorescein-labelled antigen-antibody complexes. Some T cells also develop receptors for the Fc of IgG, and these are also a feature of monocytes, macrophages and polymorphonuclear leukocytes. Monocytes and polymorphs and some B cells also bind complement component C3b. On incubation *in vitro*, some T cells develop receptors for the Fc part of IgM.

Receptors for red cells. Human thymocytes and T lymphocytes are capable of binding sheep red cells *in vitro* to form rosettes (Fig. 5.16). This is a chance finding, the significance of which is obscure.

In our experience, approximately 70 per cent of human blood lymphocytes form rosettes with sheep red cells; about 28 per cent form Fc rosettes with antibody-sensitised red cells, and of these rather less than half are B cells (positive for surface Ig) and the remainder ('null' cells) include the K-cell population (p. 152).



Fig. 5.15 The demonstration of surface Ig in human B lymphocytes by immunofluorescence, using labelled anti-Ig. Surface Ig is seen in four cells: it is forming aggregates on the cell surface, and in the upper central cell exhibits 'capping' (p. 125).

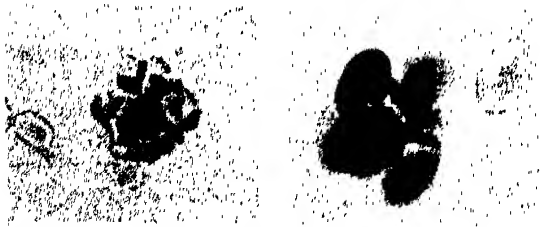


Fig. 5.16 Use of rosette formation to detect lymphocyte surface markers. **a**, a T lymphocyte binding sheep red cells. **b**, a lymphocyte binding (chicken) red cells sensitised with IgG antibody, indicating that the lymphocyte has surface receptors for the Fc of IgG.

Electron microscopy reveals differences between T and B cells, but some cells have features intermediate between the two.

The long-lived recirculating lymphocytes

In their experiments on rats, Gowans and others (p. 115) showed that when thoracic duct drainage was continued for several days, the number of small lymphocytes in thoracic duct lymph fell rapidly and the lymph nodes and spleen became partially depleted of small lymphocytes. They also showed that lymphocyte depletion was prevented by re-injecting the lymphocytes *intravenously*, and that when the cells were radio-labelled before return to the animal, most of the labelled cells reappeared in the thoracic duct lymph within the next few days. It was thus demonstrated that large numbers of small lymphocytes recirculate continuously between the blood, secondary lymphoid tissues and major lymphatics. This has since been shown to apply also to the mouse, sheep, bovines and man. Curiously, recirculation of lymphocytes is much less apparent in the pig.

Normally, about 80 per cent of the thoracic duct small lymphocytes are T cells and the rest are B cells. In thoracic duct drainage, the number of T cells removed falls rapidly, while the fall in numbers of B cells is much slower. Correspondingly, the lymph nodes, spleen and blood are depleted of T cells rapidly and of B cells slowly. In the mouse, this difference in depletion rate has been shown to be due partly to the greater rapidity of T-lymphocyte recirculation, and partly to the longer lifespan of T cells. By contrast, B lymphocytes recirculate more slowly and have a shorter lifespan.

The pathway and lifespan of recirculating T lymphocytes

The recirculating pathway of T lymphocytes has been elucidated by Parrott and de Sousa (1971) and other workers, usually by injecting intravenously lymphocytes radio-labelled with ^3H -thymidine or ^3H -uridine (which is incorporated into RNA) and following their distribution in the body by autoradiography of tissue sections or smears. Such experiments have shown that T lymphocytes in the blood gain entrance to the lymph nodes by migrating

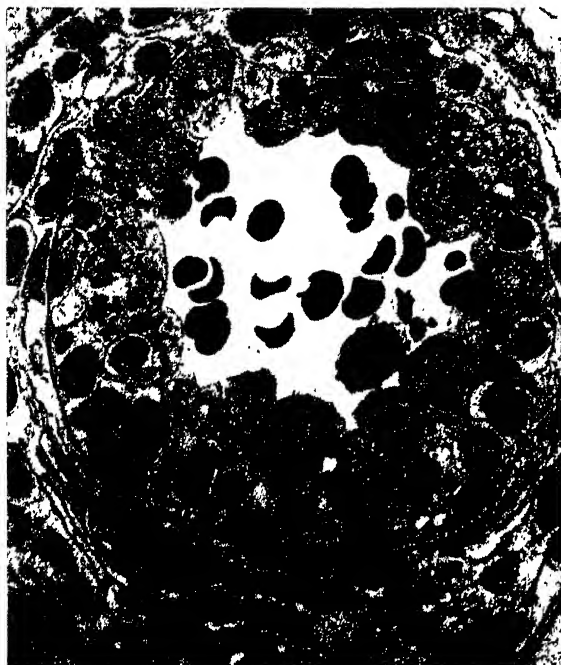
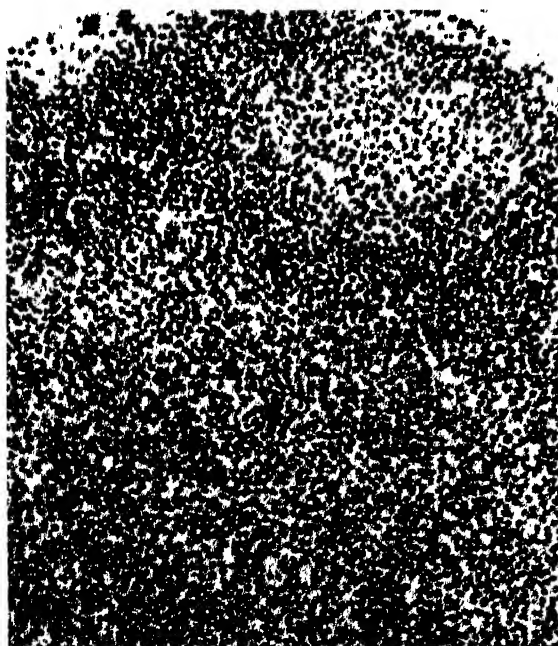


Fig. 5.17 Electron micrograph of a post-capillary venule in a Peyer's patch from a rat, showing a lymphocyte on the luminal surface, several lying between endothelial cells, and others in the surrounding sheath. $\times 1200$. (Dr. Gutta I. Schoeff.)

through the walls of post-capillary venules in the **paracortex** (also termed **deep cortex**) of the nodes. These venules have unusually tall endothelial cells, and the migrating lymphocytes pass between them (Fig. 5.17) to enter the paracortex, which contains mainly T lymphocytes and has been termed the **T-dependent zone** of lymph nodes. In neonatally thymectomised or athymic (*nu nu*) mice, these zones lack lymphocytes (Fig. 5.18) and similar depletion is produced by thoracic duct drainage or by destroying T lymphocytes by a heterologous anti-lymphocyte serum. Similar depletion of the T-dependent zones is observed in children with congenital thymic aplasia (p. 170). From the paracortex, the lymphocytes migrate to the medullary sinuses, and thus to the efferent lymphatic. They then pass via the lymphatics to the blood, thus completing the cycle (Fig. 5.19). T lymphocytes pursue a similar course through T-dependent zones of the tonsils, Peyer's patches and Malpighian bodies (white pulp) of the spleen. In the mouse, they pass through the Malpighian bodies of the spleen in about 6 hours and through the lymph nodes in about 18 hours.

T lymphocytes also leave the blood, although



a



b

Fig. 5.18 Mouse lymph nodes, showing the influence of the thymus on the histological appearances. **a**, normal lymph node, showing a cortical follicle (top right), and the deep cortex which occupies the lower two thirds of the field. **b**, lymph node from an athymic mouse: the superficial cortex shows little abnormality, but the deep cortex is almost devoid of lymphoid cells and consists largely of 'reticulum cells'. $\times 150$. (Professor D. M. V. Parrott and Dr. M. A. B. de Sousa.)

in smaller numbers, in the various non-lymphoid tissues of the body and return to the draining lymph nodes via the afferent lymphatics.

As explained on p. 119, the recirculating small T lymphocytes include memory cells derived from a previous antigenic stimulus. Recent work suggests that virgin lymphocytes supplied by the thymus do not join the pool directly.

In the mouse, the number of virgin T lymphocytes leaving the thymus is much greater than the number entering the recirculating T-lymphocyte pool. When the nuclei of thymic cortical lymphocytes are radio-labelled by injecting ^3H -thymidine directly into the thymus, labelled cells can be detected subsequently in the Malpighian bodies of the spleen, but only a few appear in the lymph nodes or thoracic duct lymph. Such cells do not, however, accumulate progressively in the spleen, and most of them apparently die within a few days of leaving the thymus. It has been suggested that, to avoid this fate, the virgin T lymphocyte must encounter an antigen to which it can respond; it then undergoes proliferation and provides effector cells and memory cells. This is supported by the fact that gnotobiotic animals (reared from birth in a germ-free state) fail to develop the recirculating T-lymphocyte pool, although lymphopoiesis in the thymus proceeds normally and some short-lived T lymphocytes (presumably virgin cells) can be detected in the spleen. They may circulate between the blood and lymph nodes, spleen, etc. during their short lifespan, but this has not been established with certainty.

It thus appears likely that the recirculating T-lymphocyte pool consists mainly of memory cells. This being so, they would be expected to be long-lived: this has been confirmed by administering injections of ^3H -thymidine over periods of several weeks and observing the rate of labelling of T lymphocytes in the recirculating pool (usually by autoradiography of cells in the thoracic duct or blood), and the duration of their persistence after stopping the injections. In mice it has been estimated that their lifespan is approximately 6 months: in rats it appears to exceed 1 year. The lifespan of human T lymphocytes is not known, but it is thought to be many years. This is based on the finding of lymphocytes with chromosomal abnormalities incompatible with successful mitosis in the blood of patients treated by radiotherapy many years before (Fig. 2.30, p. 37).

Analysis of small lymphocytes in the blood,

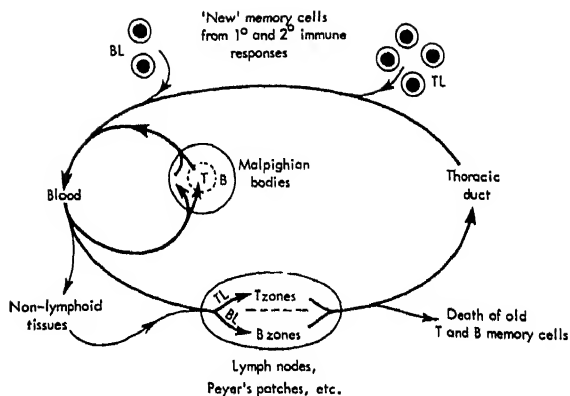


Fig. 5.19 The recirculation pathway of long-lived T lymphocytes (TL) and B lymphocytes (BL).

lymph nodes and thoracic duct has shown that 70–80 per cent of them are cells of the recirculating T-lymphocyte pool, and the degree of depletion effected by prolonged thoracic duct drainage agrees with this estimate.

A small percentage of T cells in the blood and thoracic duct are not small lymphocytes but larger 'blast' cells. These have arisen in the proliferative responses to various antigenic stimuli: experimental studies suggest that many of them settle and die in the secondary lymphoid tissues, notably those lining the gut, while a few differentiate into small lymphocytes and join the recirculating pool.

The pathway and lifespan of recirculating B lymphocytes

Most, if not all, of the small B lymphocytes in the blood, thoracic duct and secondary lymphoid tissues are recirculating cells. This has been established by experiments similar to those described above for T cells. They appear to leave the blood by the same route as T lymphocytes, i.e. via the post-capillary venules in lymph nodes, tonsils, Peyer's patches, etc., but they then come to occupy **B-dependent zones** of these secondary lymphoid tissues: in the lymph nodes they pass into and through the superficial cortex (cortical nodules—Fig. 5.30, p. 136) and medulla, and leave by the efferent lymphatic to return to the blood (Fig. 5.19). In the spleen they pass into the peripheral zone of the Malpighian bodies and presumably leave by the venous sinuses of the red pulp. These B-cell zones are depleted by thoracic duct drainage, but this takes longer than T-cell depletion because the small B lymphocytes recirculate less

actively, and their lifespan (in mice approximately 6 weeks) is distinctly shorter than that of recirculating T lymphocytes.

There are occasional large B 'blast' cells in the thoracic duct and blood: these cells have arisen in the proliferative response to various antigenic stimuli: many of them differentiate into plasma cells, some in the lamina propria of the gut where they produce IgA antibodies, others in the lymph nodes, etc. where they mostly produce antibodies of other classes.

As with the T cells, it is unlikely that virgin B lymphocytes leaving the primary lymphoid organ (in mammals the haemopoietic marrow) enter directly into the recirculating lymphocyte pool. The number of B lymphocytes leaving the marrow appears to exceed greatly the number of lymphocytes entering the recirculating pool. Like virgin T cells, many apparently die quickly, probably in the spleen, and it may be that the recirculating B lymphocytes are mainly memory cells resulting from proliferation of those virgin cells which meet an antigen to which they can respond by proliferation.

Accordingly, the recirculating T and B lymphocytes, which make up the great majority of small lymphocytes in the secondary lymphoid organs, blood and lymphatics, are likely to consist mainly of relatively long-lived memory cells.

Once antigen has been encountered, the primary immune response will have resulted in addition of specifically responsive T and B memory cells to the recirculating lymphocyte pool. On subsequent encounter with the same antigen, it is these memory cells which are responsible for the rapid and enhanced secondary antibody response and strong cell-mediated immunity (Fig. 5.20).

In both primary and secondary antibody responses, the proliferating B cells differentiate into effector cells (antibody-producing end-stage plasma cells) and B memory cells which

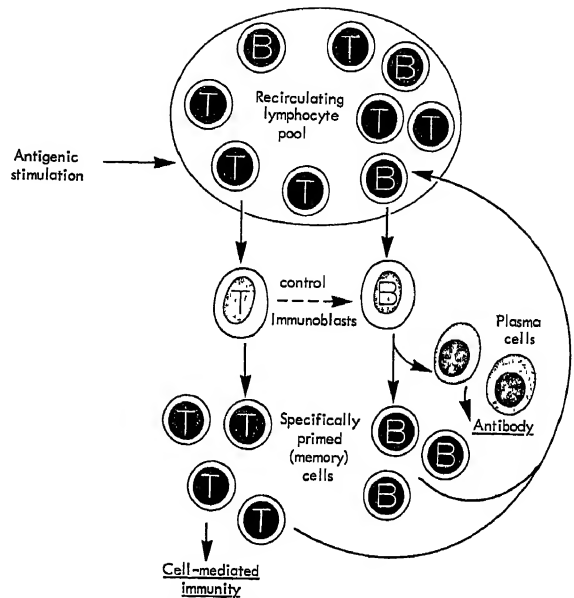


Fig. 5.20 The secondary immune response. The appropriately primed T and B memory cells in the recirculating lymphocyte pool proliferate in the secondary lymphoid tissues, providing memory cells (which re-join the recirculating pool) and effector cells. The effector B cell is the plasma cell. The effector T cells include proliferating 'blast' cells and probably also the T memory cells (see p. 119). The primary response is similar, but depends on relatively few virgin T and B lymphocytes responsive to the antigen, and so is slower and smaller than the secondary response.

join the re-circulating pool. In T-cell responses, the distinction between effector and memory cells is not so clear cut: the effector T cell is capable of producing lymphokines on reacting with the appropriate antigen, and of killing target cells (e.g. of allografts) which incorporate the antigen in their surface membrane. Both the proliferating blast cells of the cell-mediated immune response and the T memory cells which result from the response are capable of these effector functions.

The specific response of individual lymphocytes to antigenic stimulation

During their maturation in the sheltered environment of the primary lymphoid organs, individual lymphocytes differentiate in such a way that each one becomes capable of responding to only a narrow range of antigenic determinants or epitopes. The result is a continu-

ous supply of virgin lymphocytes of such considerable diversity that no matter what or how many natural or artificially-prepared foreign antigens are introduced into the body, there are likely to be some lymphocytes capable of mounting a specific immune response against

each sort of determinant on each antigenic molecule. The restricted responsiveness and diversity of lymphocytes also ensures that only a very small proportion of the available lymphocytes can respond to any one antigen, so that immune responses to many different antigens can proceed simultaneously. The development of this restricted responsiveness of individual lymphocytes comes about by changes in the cell's DNA: it occurs in both B and T lymphocytes and is heritable and irreversible, so that when a mature virgin lymphocyte meets an antigen to which it can respond, it does so by proliferation and all the cells of the resulting clone have the same restricted specificity of response (Fig. 5.21). Encounter with an antigen thus leads to an increase in the number of lymphocytes which can respond to it. Some of the cells so produced are long-lived memory lymphocytes, which, on subsequently encountering the same or a closely similar antigen, also undergo proliferation. The result is that large numbers of responsive lymphocytes are avail-

able for those antigens which are encountered frequently.

The evidence for this restricted potential of lymphocytes, and its mechanism, must now be considered. B lymphocytes have yielded up some of the secrets of their behaviour more readily than T lymphocytes, and although they are influenced by T cells, it is convenient to consider B-cell responses before T-cell responses.

The response of B lymphocytes to antigenic stimulation

Antigenic stimulation induces proliferation of B lymphocytes in the secondary lymphoid tissues. In lymph nodes, for example, this is seen in the germinal centres of the superficial cortex where responding B lymphocytes enlarge to become B 'blast' cells (B immunoblasts) with increase in both nuclear DNA and cytoplasm: the latter becomes rich in RNA and so is basophilic in its staining properties (Fig. 5.22). By successive divisions, large numbers of cells are produced, some of which differentiate into **plasmablasts**, with abundant rough endoplasmic reticulum (p. 107) and finally give rise to mature **plasma cells** which synthesise and secrete antibody (Figs. 5.3, 5.4, p. 108): others differentiate into **B-memory cells** and join the recirculating pool of long-lived small lymphocytes (Fig. 5.20).

Both virgin and memory B lymphocytes respond as described above to antigenic stimulation (Fig. 5.22). The primary response does not produce very much antibody, presumably because it depends entirely on virgin lymphocytes, relatively few of which are capable of responding to any particular antigen. The primary response does, however, provide memory cells, and these account for the much greater amount of antibody produced in the secondary response.

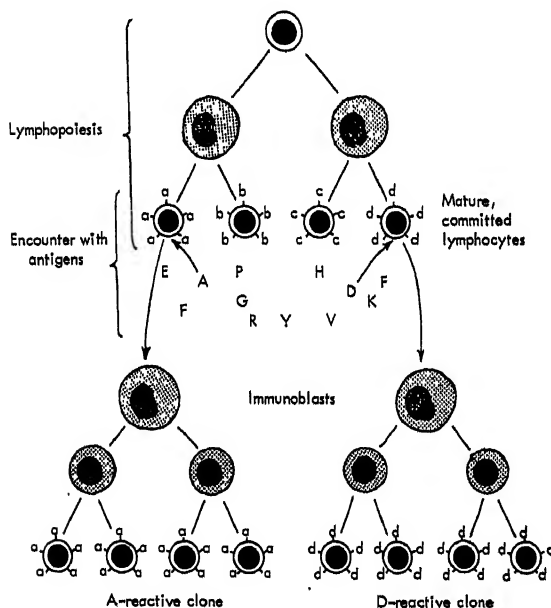


Fig. 5.21 The clonal selection theory of immune response. During lymphopoiesis, each developing T or B cell becomes committed to respond to a narrow range of antigenic determinants: this is reflected by the specificity of the antigen receptors (a, b, c, etc.) on its surface. For example, lymphocytes with hypothetical 'a' receptors can bind an antigen 'A', but not 'D' or 'E', etc. Binding of an antigen stimulates a lymphocyte to proliferate, producing a clone of lymphocytes with identical commitment.

How B lymphocytes recognise antigens

To respond to an antigen, a lymphocyte must have surface receptors capable of recognising antigen to which the cell can respond, and union of antigen with surface receptors must stimulate the response. In fact, when lymphocytes from a normal individual are incubated with an antigen which has been labelled,

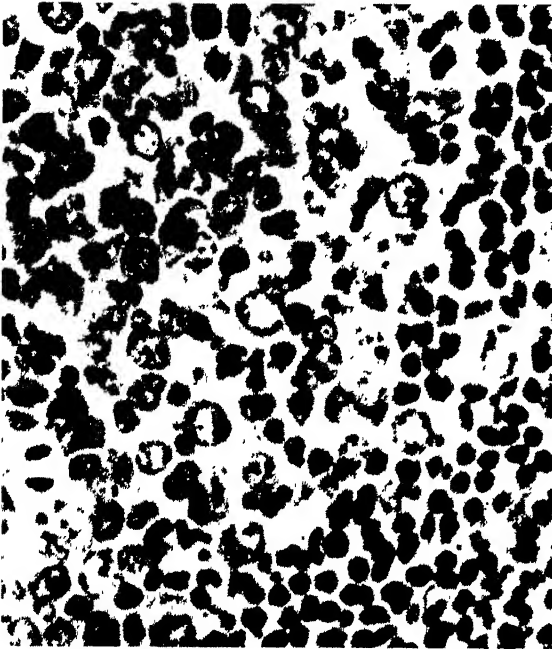


Fig. 5.22 Segment of a germinal centre showing proliferating B 'blast' cells with large pale nuclei and abundant cytoplasm. The centre is surrounded by tightly packed small lymphocytes. $\times 850$.

for example with ^{131}I or fluorescein, the labelled antigens can be demonstrated to bind to the surface of very small proportions of both T and B lymphocytes. If the donor has previously encountered that antigen, the number of antigen-binding lymphocytes is increased, but is still only a very small proportion of the total.

The B lymphocyte has on its surface molecules of immunoglobulin (SIg), which float in the lipid of the cell membrane with the Fab end of the molecules projecting from it, and there is now no doubt that these act as the specific antigen receptors on B cells. Some of the evidence for this is summarised below.

Firstly, treatment of B lymphocytes with an anti-immunoglobulin (anti-Ig) inhibits antigen binding.

Secondly, evidence has been provided by the 'capping' phenomenon. When B lymphocytes are treated with a labelled antigen, the SIg molecules of those cells which bind the antigen become aggregated into clumps and finally in a single mass or 'cap' at one part of the surface: the effect is the same as that induced by treating B cells with anti-immunoglobulin (Fig. 5.15, p. 120). When capping is induced by antigen, the cap can be shown to contain *all* the antigen bound to the cell surface and *all* the SIg molecules, thus demonstrating that all the SIg molecules react with the antigen.

The significance of antigen binding by the surface Ig of B cells

Those B cells which can bind a particular antigen are responsible for the antibody response to that antigen. This has been demonstrated by passing lymphocytes through a column of inert material coated with antigen. The cells which bind the antigen are retained and are thus separated from non-binding lymphocytes. The separated cells may be tested for antibody response either *in vitro* or by administering them to an animal whose own lymphocytes have been destroyed (e.g. by x-irradiation) and then challenging the animal with the same and other antigens. Such experiments have shown that the antigen-binding cells can mount an antibody response to the same antigen. The non-antigen-binding population does not respond to the same antigen, but responds normally to other antigens. These findings suggest that the SIg molecules of B lymphocytes represent a sample of the specific antibody which that cell and its descendants can synthesise and secrete. This is further supported by the observations outlined below on individual clones of proliferating cells.

The sequences of amino acids in the variable regions of the light and heavy chains of the immunoglobulin (Ig) molecule, including the antigen-combining sites (Fig. 5.1, p. 106) constitute what is known as the **idiotype** of the molecule and are responsible for its specificity as an antibody. The idiotype of the molecule might be expected also to exhibit specific *antigenic* properties, but this is difficult to establish because normal Ig, e.g. in serum, is a mixture of enormous numbers of antibodies, each with its own idiotype. However, some B-cell lymphoid tumours, for example myeloma, differentiate into plasma cells and secrete molecules of Ig which are identical for any one such tumour. This can only be explained by concluding that the Ig is **monoclonal**, i.e. that the tumour arises from clonal proliferation of a B cell already committed to production of Ig of a particular idiotype. *It follows that the commitment of a B cell to production of Ig of a particular idiotype, i.e. of a particular antibody, is transmitted to its descendants.*

In some of these Ig-secreting tumours, the tumour cells resemble normal B lymphocytes in having surface Ig (SIg), and this has provided an opportunity to compare the secretory and

surface Ig of a B-cell clone. When the monoclonal Ig (obtainable from the patient's serum) is injected into a rabbit, the antiserum contains antibodies to the variable regions, and these **anti-idiotypic antibodies** react with, and cause complete capping of, the SIg of the tumour cells (p. 120). This is strong evidence that B-cell SIg has the same idio type as the antibody secreted by that cell and its descendants. There is now evidence that this applies also to normal (i.e. non-neoplastic) B cells, for by injecting mice of a selected inbred strain with a polysaccharide of limited antigenicity, relatively few lymphocytes are stimulated to clonal proliferation and antibody production, and it has been possible to demonstrate that an anti-idiotypic antibody reacts with both the free antibody and the SIg of the cell clone responsible for it. Accordingly, it may be concluded that *the surface Ig of normal B lymphocytes represents a sample of the specific antibody which that cell or its descendants can produce in response to an antigenic stimulus*.

How B lymphocytes develop specific responsiveness and diversity

From the preceding sections it is apparent that, at some stage in its development, each B lymphocyte becomes committed to producing antibody of a particular specificity. Since the specificity of antibody depends on the sequence of amino acids in the variable regions of the polypeptide chains, it is a reflection of the sequences of the bases in the DNA of the genes coding for immunoglobulin synthesis. Commitment of the lymphocyte thus implies that it becomes restricted to coding for only one particular sequence for the variable regions of the heavy chain and one sequence for the variable regions of the light chain. Since all the B cells of a proliferating clone are committed to produce the same antibody, commitment must be heritable and irreversible through many cell generations. But individual lymphocytes become specifically committed to produce *different* antibodies, thus providing the great diversity of the antibody response.

We must therefore conclude that either (a) the genome of the individual contains large numbers of alternative genes for the variable parts of the Ig chains (V genes), and that commitment involves the restriction to coding for one light and one heavy chain V gene, or (b) the genome of the individual contains relatively few V genes, and different sequences of bases develop by somatic mutation during early lymphopoiesis (Jerne, 1971). It has not been established with certainty which of these two theories of diversity is the correct one, but present evidence favours the former (which is assumed in this account), i.e. that the genome of the individual contains sufficient V genes to account for the whole repertoire of antibodies which the individual is capable of producing (**the germ line theory**), and each B lymphocyte becomes committed to specific antibody production by restriction of coding to one V gene for each Ig chain.

It is known that each Ig chain is synthesised as a single unit, by translation of messenger RNA carrying the code for the constant and variable regions of the whole chain. It is now believed that the gene selection responsible for lymphocytic commitment and diversity is brought about by the mechanism illustrated for heavy chain synthesis in Fig. 5.23. The heavy chain V genes (VH genes) form a continuous series and by formation and exclusion of a loop of random length, a particular VH gene segment is brought into apposition with the gene segment coding for the constant part of the heavy chain (C μ , C γ , etc.). Once this has occurred, the adjacent V and C genes act as a single cistron, and the sequence of amino acids in the variable regions of the heavy chain of Ig produced by that cell and its descendants is irrevocably decided.

Recently this mechanism of rearrangement of gene segments has been shown to operate in the formation of Ig light chains, in which three DNA segments come together to form the whole chain (Fig. 5.24). The antibody specificity is decided by random combination of a VL and a 'J' gene segment,*with which the C segment also combines to complete the cistron.† In the mouse κ chains there are approximately 1000 VL gene segments and 10 'J' segments, giving a repertoire of 10^4 amino-acid se-

*The sequences of DNA which encode the antigen-combining component of light chains have recently been shown to be divided into VL gene segments and J segments. The J segments are quite unrelated to the J chains which bind together monomeric IgA and IgM to form polymers.

†Separation of segments of a single gene on the DNA strand of the chromosome has been shown also in relation to production of ovalbumin and other proteins and may apply to eukaryotic proteins in general.

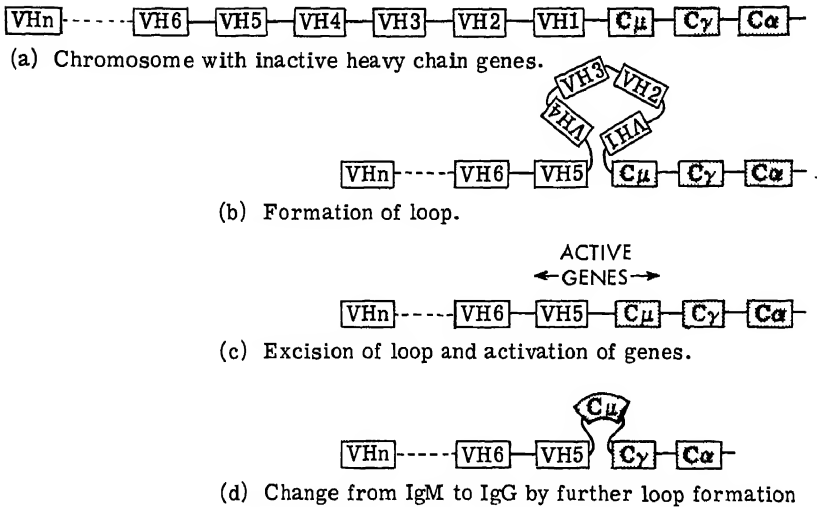


Fig. 5.23 Model of chromosome with genes for the variable region (VH1, VH2 ... VHn) and constant region (Cμ, Cγ, Cα) of heavy chains, showing selection of a single VH gene by formation and removal of a DNA loop. In this example, the cell becomes committed to synthesis of heavy chains with a variable region containing the amino-acid sequence determined by VH5. At first IgM is synthesised, but removal of a loop containing Cμ leads to subsequent synthesis of IgG with the same specificity (determined by VH5).

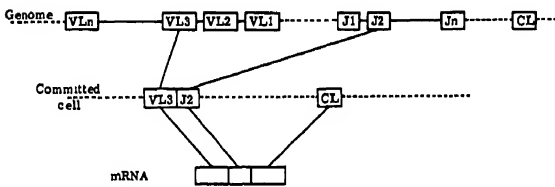


Fig. 5.24 Gene segments involved in coding for immunoglobulin light chain. The variable region is determined by selection of one of approximately 1000 VL gene segments; these are separated from 10 or so J gene segments which also determine part of the variable regions. The J segments are separated from the (C) segment which codes for the constant region. Commitment of the cell depends on random apposition (by looping) of one VL and one J segment. The superfluous bases between the J and C segments are transcribed but the corresponding part of the messenger RNA (mRNA) is excised, leaving a continuous mRNA strand coding for the whole light chain. (Modified from Williamson, 1979).

quences. The specificity of antibody depends on the variable regions of light and heavy chains, and if we assume a similar variability (10^4) for the mouse heavy chain, this gives a total repertoire of 10^8 different antibodies, which agrees fairly closely with most estimates based on analyses of the antibodies which develop in response to one particular antigenic determinant. (See Williamson, 1979).

The theory of random selection of gene segments coding for the variable parts of Ig which

form the specific antigen receptors explains not only the specificity of immune responses but also the multiplicity of antibodies which, as mentioned above, are produced, even in response to a simple antigen with only one type of determinant. The antigen stimulates all those B lymphocytes whose specific surface receptors can bind it sufficiently firmly and each of these cells gives rise to a clone of plasma cells producing its own distinctive antibody molecules. The result is a mixture of diverse antibodies, some of which can bind the antigen more firmly than others, but the total effect of which is exquisite specificity. If the antigen is administered repeatedly in small doses it is taken up mainly by those lymphocytes which can bind it most firmly and as these proliferate the result predicted would be a progressive increase in the avidity (p. 107) of the antibody, which is indeed what happens.

It is now widely accepted that commitment of B lymphocytes occurs spontaneously, probably during lymphopoiesis in the sheltered environment of the haemopoietic tissue, and is not dependent on encounter with an antigen. This view appears to have been held as early as 1908 by Ehrlich, but was for long ignored in favour of the *instructive theory* which postulated that antigenic material was taken up by antibody-producing cells and instructed the cell to make antibody by acting as a template on

which the antibody was moulded. Ehrlich's view was revived by Jerne (1955) and by Burnet (1959) as the **clonal or cell selection theory** (Fig. 5.21, p. 124). Once it was established that antibody specificity was determined by the sequence of amino acids in the Ig chains, the instructive theory became untenable, and the evidence in favour of the clonal selection theory, some of which has been outlined in the preceding account, is now overwhelming.

The classes of antibody produced by B cells

The class of Ig synthesised by a B cell will depend on which C-gene segment is linked to the selected VH gene, and change of production from one class to another, e.g. IgM to IgG, probably involves similar looping and exclusion of C-gene segments as shown in Fig. 5.23. It is noteworthy that the class of Ig synthesised (embodied in the Fc portion of the heavy chains) may change during the development of a clone of antibody-producing cells, but the light chain type and idiotype do not change.

During maturation in the haemopoietic marrow, B cells pass through a stage in which they synthesise and secrete small amounts of monomeric IgM. This is followed by the appearance of monomeric IgM on the surface membrane. In the primary immune response, the proliferating B lymphocytes have mainly surface IgM, and differentiate into plasma cells which secrete IgM antibody. Later in the response some of the proliferating B cells develop into plasma cells which secrete IgG and other classes of antibody. Memory B lymphocytes (long-lived small B lymphocytes of the recirculating pool—p. 122) have mainly IgD on their surface membrane, but give rise predominantly to plasma cells which secrete IgG or IgA antibodies, as observed in the secondary response.

The response of T cells to antigenic stimulation

Antigenic stimulation induces proliferation of responsive T lymphocytes: the proliferating cells (like responding B cells) become enlarged and acquire abundant basophilic (RNA-rich) cytoplasm (Fig. 5.25): they are termed **T 'blast' cells** or **T immunoblasts**. They pass through a succession of divisions during which they

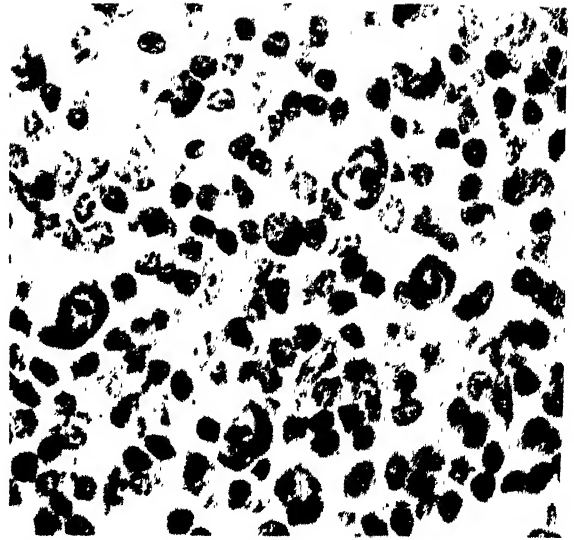


Fig. 5.25 Large lymphoid cells with enlarged nucleus and pyroninophilic cytoplasm (*immunoblasts*) in a human lymph node removed 6 days after a skin allograft. Methyl green pyronin stain. $\times 480$. (Professor D. M. V. Parrott and Dr. M. A. B. de Sousa.)

differentiate into small T lymphocytes capable of binding the antigen which has induced their production. T-cell proliferation takes place in the T-dependent zones of the secondary lymphoid tissues (p. 121).

Like B lymphocytes, T lymphocytes become committed to respond to a narrow range of antigenic stimuli, and this occurs without antigenic stimulation, probably during lymphocytic maturation in the thymus. Another similarity is that the T lymphocyte recognises antigens by means of surface receptors, but the nature of these is uncertain. The T cell has not been shown to synthesise Ig at any stage of its development, and conventional Ig cannot be detected on its surface.* Yet T cells show specific antigen-binding properties, and the development of cell-mediated immunity depends on such antigen binding. It still remains possible, however, that the T-cell antigen receptor consists of amino-acid sequences which are the products of VH and VL genes, i.e. the variable regions of Ig molecules, for these are not detectable by conventional anti-Ig. Support for this view has been provided by the following experiments.

By administering to inbred mice or rats an antigen which in these animals stimulates only a few T and B

*When blood is withdrawn and allowed to cool, plasma IgG adheres to the surface of T lymphocytes, but dissociates readily at 37 °C.

lymphocytes to clonal proliferation, it has been possible to isolate antibody of a particular idiotype (p. 125) and to prepare an anti-idiotype antibody. In such experiments it has been shown that the anti-idiotype antibody reacts with the receptors of both T- and B-cell clones, which suggests strongly that they are identical, and since the B-cell receptors are known to be the variable ends of Ig molecules, it seems likely that the T-cell receptors are also of this nature. Alternatively, T-cell receptors could be determined by a set of genes analogous to, but distinct from, the V genes which code for B cells: possible contenders are the *Ir* (immune response) genes (p. 167) which somehow influence T-cell-dependent responses.

The remainder of the T-cell receptor molecule, corresponding to the constant regions of Ig, has not been shown to be any of the known classes of Ig (μ , γ , etc.), and its nature is unknown. The C genes of B cells thus appear to be represented in T cells by another gene, or series of genes, of unknown nature. Accepting this assumption, the commitment of T lymphocytes can be explained by the same looping and exclusion mechanisms illustrated in Fig. 5.23.

In summary, T lymphocytes owe their specific responsiveness to antigen to the presence on their surface of antigen-specific receptors. Unlike the specific (SIg) receptors of B cells, the T-cell receptors are not whole Ig molecules, but in some instances their antigen-binding sites have been shown to be similar (and possibly identical) to the idiotypes of SIg. The part of the T-cell receptor corresponding to the Fc of SIg is, however, of unknown nature.

Control of the immune response

So far, this account might suggest that antibody and cell-mediated immune responses arise when B or T lymphocytes respectively encounter an antigen to which they can respond. In fact, immune responses are far more complicated than this, and although antigenic stimulation is obviously of central importance in triggering them off, the nature of the antigen, its molecular size and density and variety of determinant sites are only some of the factors concerned. Account must be taken of many other factors which can enhance, suppress or modify immune responses. Some of these are outlined briefly below.

T-dependent and T-independent antibody responses. Animals deficient in T cells (e.g. 'B' mice, p. 117) are unable to mount antibody responses to many antigens, including foreign

proteins. B cells require the co-operation of T cells to produce antibodies to such antigens, which accordingly are called **thymus-dependent antigens**. Some antigens, however, do stimulate 'B' mice to antibody production: these so-called **thymus-independent antigens** are usually of very high molecular weight, with large numbers of one or more particular antigenic determinants, e.g. bacterial lipopolysaccharide, pneumococcal capsular polysaccharide or artificially prepared polymers. It appears that, by binding to, and so linking up, large numbers of the surface Ig receptors on a B cell, such antigens can provide the necessary stimulus for clonal proliferation and antibody production (Fig. 5.26). By contrast, smaller molecules, or globular proteins which display a variety of antigenic determinants without a high concentration of any one determinant, would be expected to bind less strongly to B cells and to occupy fewer surface receptor sites: this may be why they cannot stimulate an antibody response without the help of T cells.

The helper function of T cells in antibody responses has been elucidated by complexing a hapten (e.g. dinitrophenyl) to a foreign carrier protein. Using this system, Mitchison (1971) and others have shown that an antibody response to the hapten is dependent on the development of cell-mediated immunity to the carrier protein. It seems likely that the binding of molecules of the protein to primed T lymphocytes renders the hapten-protein complex

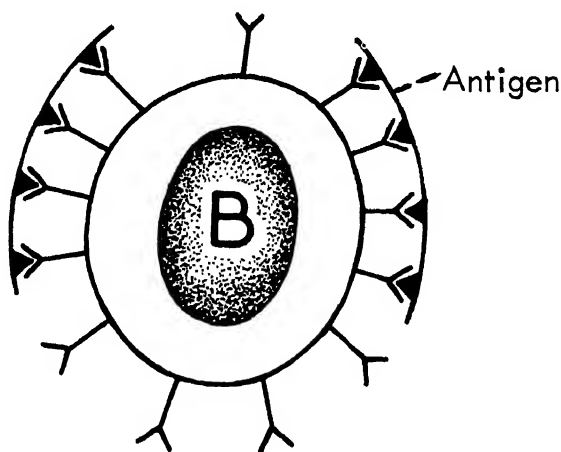


Fig. 5.26 T-independent antibody response induced by an antigen with large numbers of a particular type of antigenic determinant (epitope) which effectively links the surface Ig antigen receptors of an appropriate B cell and stimulates it to respond.

immunogenic to B cells, perhaps by presenting the hapten to B cells in such a way that it effectively links together the B cell receptors (Fig. 5.27b). Alternatively, it could be that the T cell binding the protein carrier transmits an additional stimulatory signal to the B cell binding the hapten. Some evidence for this 'second signal' hypothesis is provided by the demonstration that T cells binding and responding to a quite separate antigen (or to mitogens—p.

119) can be of some help to the B cell in its response (Fig. 5.27d). It is also possible that the T cell responding to the protein carrier transfers receptor sites to macrophages, and thus confers on them specific helper capacity (Fig. 5.27c).

Whatever the mechanism of T-cell co-operation in antibody production, it has been shown to be a property of memory T cells and, as indicated above, can be antigen-specific or non-specific. Specific T-cell co-operation is essential for optimal primary and secondary antibody responses to most protein antigens, and even those immune responses which can occur without T-cell help are usually limited mainly to production of antibodies of IgM class. The co-operation of T cells appears to be necessary for the production of large amounts of IgG antibody, as in the secondary response. Indeed, there is evidence that the formation of B-memory lymphocytes is less dependent on T-cell help than is the production of antibody-producing plasma cells. The degree of helper T-cell activity in the response of inbred strains of mice to highly artificial T-dependent antigens has been shown to involve the so-called Ir (immune response) genes which lie within the major histocompatibility gene complex (p. 167). While it is possible that Ir genes code for specific antigen receptors on the surface of T cells, there is increasing evidence that many genes of the major histocompatibility complex are concerned with the co-operation of T cells with B cells and with macrophages (Munro and Waldmann, 1978).

Suppressor T cells. In addition to T cells which co-operate in the antibody response, it is

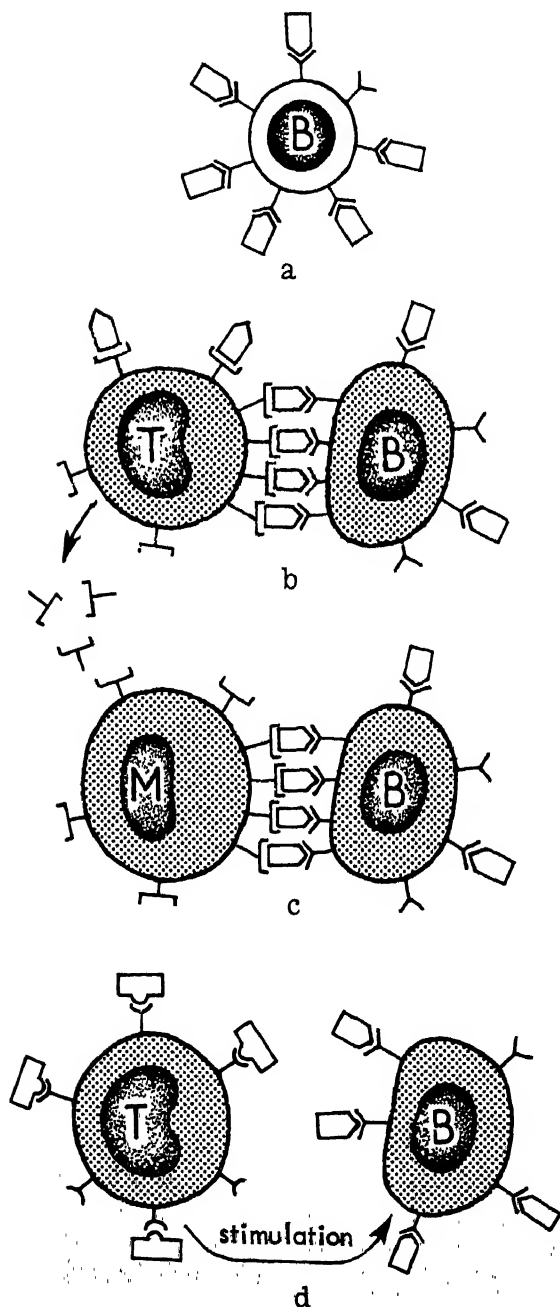


Fig. 5.27 Possible mechanisms of co-operation of T cells in antibody production by B cells. In **a**, a B lymphocyte has bound an antigen by a particular type of determinant, but is not stimulated by it because it has not cross-linked the specific antigen receptors. If a T cell also binds the same antigen by a different determinant, as in **b**, it may effectively cross-link the B-lymphocyte receptors and thus stimulate it (indicated here by 'blast' transformation of the B cell). Alternately the T-cell antigen receptors may be transferred to macrophages **c** which then become capable of co-operating in B-cell stimulation. Lastly, a T cell which has bound an independent antigen, as in **d** can also provide a stimulus to the B lymphocyte. In all these forms of co-operation, the stimulating T cell has itself been stimulated to 'blast' transformation, and indeed the last form of co-operation can be provided by a T cell which has been stimulated non-specifically, e.g. by phytohaemagglutinin.

now known that subsets of T cells have a suppressive effect on B cells. This explains why the antibody response to thymus-independent antigens is often greater in T-deficient animals. In lepromatous leprosy and some other chronic infections, very high titres of antibodies to the causal agent are commonly observed and may well be due to the lack of suppressor T-cell control, for failure of cell-mediated immunity (and therefore of T-cell function) to the micro-organism is a feature in such cases.

The helper and suppressor activities of T cells are complex, and it seems likely that the T cells responsible are distinct sub-sets rather than stages in the differentiation of one series of T cells. The complexity is increased by recent evidence that helper T cells (and probably suppressor T cells) exert some control over T cell, as well as B cell, immune responses.

Antibodies. The immune response to an antigen can be partially or even completely inhibited by injection of the corresponding antibody either before or shortly after administration of the antigen. Part of the inhibitory effect is due to rapid phagocytosis and destruction of much of the antigen following its union with the injected antibody. There is, however, evidence that antibody can exert an additional inhibitory effect, which is not due to destruction of antigen, and may involve receptors on suppressor T cells for the Fc of IgG. The inhibitory effects of antibodies are of practical importance in prophylactic immunisation of infants, for maternal IgG antibodies cross the human placenta and may interfere with the response to vaccines during the first few months of infancy. The inhibitory effects of antibody are now used extensively to prevent immunisation of the rhesus-negative mother by a rhesus-positive fetus (p. 151).

Anti-idiotype antibodies. As explained earlier (p. 104), most antigens are of exceedingly complex structure, and each stimulates large numbers of different T and B cells to clonal proliferation. In consequence, the antibody produced is really a mixture of large numbers of different antibodies, each the product of a clone of B lymphocytes. Even antigens with a single type of determinant usually stimulate clonal proliferation of many different lymphocytes, and a mixture of antibodies results (p. 127). In spite of this heterogeneity of the antibody response, there is evidence that the variable parts

of the antibody molecules resulting from any particular clonal proliferation may act as *antigens* in the host and stimulate the development of anti-idiotype antibodies (p. 126). These can bind to the antigen receptors of the corresponding B- and T-cell clones (p. 129) or of the corresponding antibody molecules, and in so doing can exert a regulatory effect on the cells of the clones. Depending on various factors, they may either enhance or suppress further clonal proliferation and differentiation, and thus exert some control over the immune response. Such control does not stop at this level, for there is now evidence that the anti-idiotype antibodies also act as antigens and may stimulate the development of further anti-idiotype antibodies (anti-anti-idiotypes), and so on.

It is thus becoming increasingly apparent that immune responses, like many other biological processes, are not regulated by a simple feedback mechanism, but by a complexity of factors, the outcome being determined by the total effect of all the stimulatory and inhibitory factors involved.

*In conclusion, the production by B cells of antibodies to many antigens is dependent on the co-operation of T cells. Such help is provided optimally by T cells which have responded to the same antigen. There is evidence that T cells may also suppress the immune response of both B and T cells, and that they play a major role in controlling the intensity of immune responses. Antibody itself is also capable of suppressing the immune response to the corresponding antigen, and anti-idiotype antibodies can both enhance and inhibit clonal proliferation of both T and B cells with receptors of the corresponding idio-*type*.*

Transfer factor

Since 1948, Lawrence has reported investigations which suggest that cell-free extracts of human leukocytes can transfer cell-mediated immunity to non-immunised recipients, and that the responsible agent, transfer factor, has a molecular weight of about 3000 (see Lawrence, 1969). Attempts to demonstrate transfer factor in experimental animals have, in general, been disappointing, but it must be appreciated that, in general, man and other primates show much stronger cell-mediated immune responses and DHS reactions than do lower animals.

Recently, a number of reports have appeared on the therapeutic use of transfer factor in the treatment of patients with resistant infections due to immunodeficiencies, and with various diseases of obscure causation (e.g. sarcoidosis, connective tissue diseases and tumours). The de-

gree of success achieved has varied greatly, and it is difficult to draw conclusions. The existence of human transfer factor which enhances immune responses is no longer in doubt, although its mode of action is obscure and its antigen-specificity has not been widely accepted.

Acquired immunological tolerance

In some circumstances, exposure to antigen does not result in an immune response, but in the development of unresponsiveness (tolerance) of the individual to that particular antigen, although responses to other antigens remain normal. In addition to true or 'classical' tolerance, suppressor T cells and antibody are both capable of inhibiting specific immune responses (see above), and thus may bring about a state resembling true tolerance.

Classical immunological tolerance. When living cells or tissues are exchanged between genetically dissimilar individuals, the host develops an immune response which results in destruction of the transplanted cells. Yet in 1945 Owen, an American veterinary surgeon, detected red cells of two distinct groups in some bovines, and noted that such animals were always derived from a twin pregnancy. Apparently there is sometimes placental vascular anastomosis between dizygotic twin cattle, with consequent admixture of their blood during early fetal life. Circulating haemopoietic stem cells from each fetus settle in the other and remain functional so that each twin subsequently produces red cells, etc. from its own and from its twin's stem cells.

This observation—that allogeneic cells introduced into the embryo were not rejected by the host—led Burnet and Fenner (1949) to postulate that antigenic challenge during fetal life could result in specific unresponsiveness rather than an immune response. Burnet failed to substantiate this because of an unlucky choice of antigen/host combination. It was, however, subsequently confirmed by Medawar and his co-workers, who showed that injection of living cells from a mouse of one inbred strain (say Y) into a *neonatal* mouse of another inbred strain (X) induced immune unresponsiveness to cells of the donor strain. Thus when the treated

mouse matured, it failed to reject a skin graft from a Y mouse, although capable of rejecting normally skin from a mouse of an unrelated strain (Z) (Fig. 5.28). The likely explanation of this form of 'classical' tolerance is that, on encountering an antigen to which it can respond, the immature lymphoid cell is either destroyed

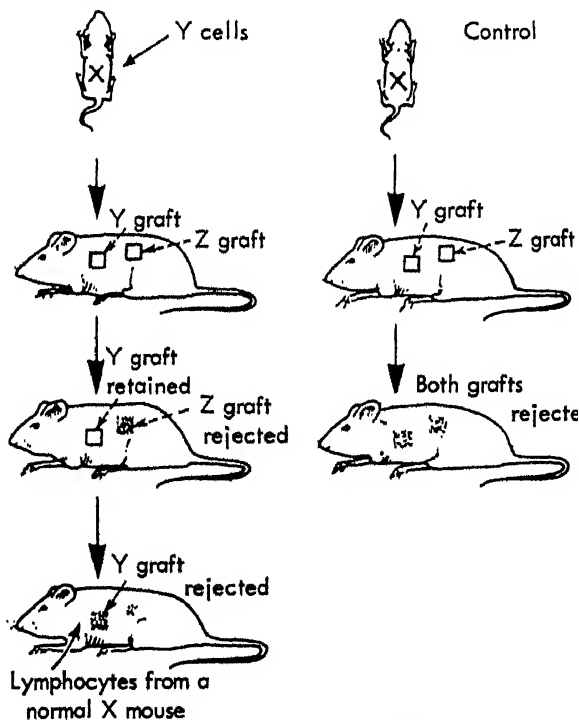


Fig. 5.28 Immunological tolerance to allogeneic cells. A neonatal 'X' strain mouse (top left) injected with strain 'Y' cells becomes tolerant to 'Y' 'transplant' antigens, as shown by retention of a subsequent 'Y' skin graft. It rejects an unrelated ('Z' strain) graft normally. Injection of lymphocytes from a normal 'X' mouse results in rejection of the 'Y' graft. (In practice, injection of lymphocytes or haemopoietic cells is commonly used to induce tolerance in such experiments: this raises the complication of the graft versus host reaction (p. 167) which, for the sake of simplicity, has been ignored in this illustration.)

or rendered irreversibly unresponsive. This is supported by the additional finding that when the tolerant X mouse bearing a Y skin graft is injected with lymphocytes from a normal adult X mouse, the Y graft is rejected, i.e. tolerance is abolished.

Tolerance may be induced also to foreign proteins by injecting them into the fetal or neonatal animal, but to maintain such tolerance it is necessary to administer repeated injections of the antigen. Thus classical tolerance is maintained only so long as the antigen persists in the host. As an explanation, it is suggested that, when lymphocytes become committed to recognise a particular antigen, they pass through an immature stage during which encounter with the antigen promotes their death or non-responsiveness: if antigen is continuously present it will 'catch' potentially responsive lymphocytes at this stage and tolerance will persist. Mature lymphocytes are more likely to mount an immune response on encountering the antigen. It follows that widely spaced pulses of antigen should favour an immune response, as indeed they do.

Although tolerance is most readily induced in immature animals, adults may also be rendered tolerant and Mitchison (1967) showed that this could be achieved by administering repeated injections of either very low or very high doses of antigen to adult mice. Intermediate doses induced an immune response. The explanation is that tolerance is readily induced in T lymphocytes by very small or very large doses of antigen, while B-lymphocyte tolerance is induced only by very large doses of antigen (Fig. 5.29). It follows that high dosage suppresses both the T-cell response (cell-mediated immunity) and the B-cell response (antibody production). If the antigen is T-dependent (p. 129), as in Mitchison's experiment, then low dosage, by inducing T-cell tolerance, will inhibit also the antibody response. For T-independent antigens, however, low dosage will induce suppression of cell-mediated immunity but there will still be antibody production: this is sometimes termed **split tolerance** or **immune deviation**.

In fact, the situation is more complex than stated above, for any particular antigenic determinant will be bound by many lymphocytes with different specific receptors (p. 127). Some of these cells will have receptors which make a 'good fit' with the antigen

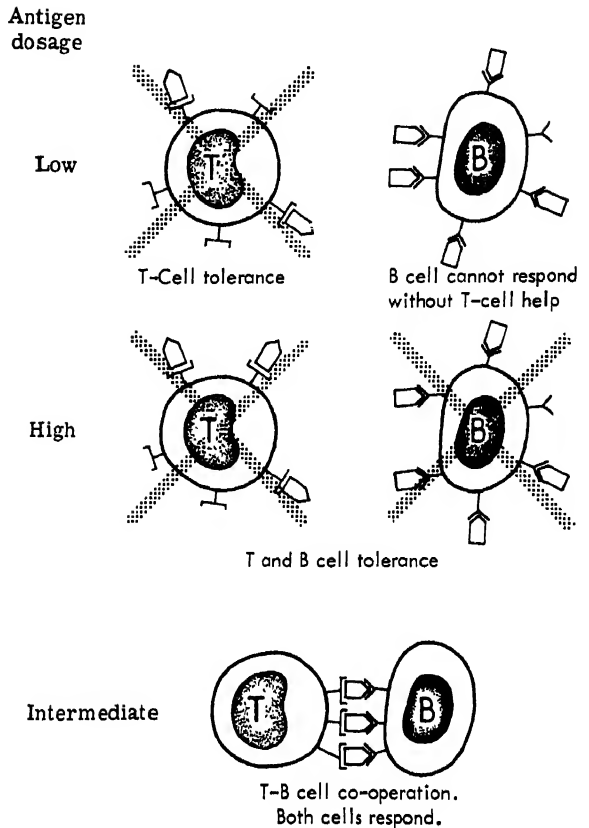


Fig. 5.29 The induction of immunological tolerance to a 'thymic-dependent' antigen by administration of very small or very large amounts of the antigen.

and bind it firmly; others will bind it less well because it fits their receptors poorly. These factors will influence not only the amount of antibody produced, but also its avidity. A high concentration of antigen which induces non-responsiveness of strongly-binding cells may stimulate an immune response in those binding the antigen less firmly, with consequent production of a relatively small amount of antibody of poor avidity. The same considerations apply to T lymphocytes, and accordingly tolerance is not an all-or-nothing phenomenon. Any antigenic stimulus is likely, in fact, to induce tolerance in some potentially responsive lymphocytes and an immune response in others.

Well-established tolerance to an antigen can sometimes be broken down by administering the antigen incorporated in Freund's adjuvant (p. 114) or by administering a closely related cross-reacting antigen. The mode of action of Freund's adjuvant is uncertain, but abolition of tolerance by a closely related antigen is probably explained by Mitchison's work with hapten bound to a protein carrier. As already noted (p. 129), he showed that antibody production to the hapten is dependent on a T-cell response to antigenic determinants of the carrier pro-

tein. If T-cell tolerance to the carrier is induced by low dosage of the hapten-carrier complex, then, as explained above, the antibody response will also be suppressed. If now the same hapten coupled to another protein is injected into the animal, T cells will respond to the new carrier protein and in doing so will provide the co-operation needed by B cells to respond to the hapten. Native proteins each possess various antigenic determinants and by regarding any particular type of determinant as equivalent to the hapten in Mitchison's experiments, we can explain the breakdown of B-cell tolerance by administration of a closely related cross-reacting antigen.

Conditions resembling classical tolerance. So far in this account, tolerance has been explained on the basis of a direct and permanent suppression of T cells, and sometimes of B cells also, by a particular antigenic stimulus. However, it was shown by Gershon and Kondo (1971) that the thoracic duct lymphocytes of a mouse rendered tolerant to a particular antigen could confer tolerance on a normal animal of the same strain. This is now regarded as due to the development of suppressor T cells capable of inhibiting T- and B-cell responses to a particular antigen. A form of unresponsiveness to an antigen can also be induced by administration of the corresponding antibody shortly before or after antigenic stimulation (p. 131).

The significance of immunological tolerance. The capacity of the immunity system to develop immune responses to countless antigens carries with it the danger of responding similarly to one's own normal body constituents, i.e. by auto-immunisation. The ease with which classical tolerance can be induced to specific antigens during fetal and neonatal life is believed to be the major safeguard against auto-immunisation, but it is not always effective and some diseases are attributable to auto-immunity (p. 161). Also B cells with surface Ig

capable of binding self-constituents, e.g. thyroglobulin or DNA, have been detected in the blood of some normal individuals, indicating a capacity for auto-antibody formation; this is apparently held in check in most individuals either by lack of T-helper cells, or by a dominating influence of suppressor T cells, and there is evidence that deficiency of the latter is a feature of auto-immune diseases.

As described above, tolerance may also develop towards the antigens of *foreign* cells or proteins introduced into the animal during fetal or neonatal life, and this can apply also to micro-organisms, with consequent persistence of infection, e.g. lymphocytic choriomeningitis of mice (p. 183).

The fact that tolerance to foreign antigens can be induced throughout life offers a promising approach to the therapeutic transplantation of foreign tissues, for the induction of specific tolerance to antigens of the donor tissue would obviously be preferable to the use of drugs which bring about a general depression of the host's immunity system to all antigens, including those of pathogenic micro-organisms. In fact, administration of such drugs (e.g. azathioprine, glucocorticoids, etc.) together with an antigen facilitates the induction of tolerance in the adult, and this is probably of importance in clinical renal transplantation, in which the dosage of immunosuppressive drugs can be gradually reduced without rejection of the kidney. Some form of tolerance develops to the alloantigens of the transplant: this may either be 'classical' tolerance, as described above, or it may be due to production of specific suppressor T cells or of 'blocking' or 'enhancing' antibody, which suppresses further immune responses, including cell-mediated immunity, to the transplant antigens (p. 167).

Macrophages and the immune response

When antigenic foreign macromolecular or finely particulate material penetrates into the body, much of it is engulfed by macrophages. Most of the ingested antigen is digested within phagolysosomes and destroyed, but some antigenic material becomes bound to the plasma

membrane and a small proportion persists within the macrophage where it is protected in some unknown way from digestion and is slowly secreted. Both forms of antigenic material—surface and secretory—persist for at least 2 weeks in immunogenic form (see below).

Macrophages can bind and engulf antigen without the assistance of antibody, but if IgG antibody is present (from previous immunisation) and combines with the antigen to form immune complexes, binding is greatly enhanced because macrophages, like polymorphs and some lymphocytes, have surface receptors for the Fc of the IgG in the complexes. Macrophages do not have receptors for IgM, but when IgM antibody reacts with antigen, complement is activated and the binding of the antigen-antibody complexes is then enhanced by surface receptors for the C3b component of complement (p. 181). Some sub-classes of IgG antibody are cytophilic for macrophages, i.e. they bind to the macrophage surface in the absence of antigen, and will enhance the binding of antigen subsequently encountered. These mechanisms of antigen binding are summarised in Fig. 7.2, p. 180.

The immunogenicity of macrophage-associated antigen. When living macrophages are incubated with an antigen *in vitro* and washed to remove free antigen, they induce an immune response when injected into a syngeneic animal. If the antigen is weak, i.e. induces a poor immune response when injected alone, prior processing by macrophages considerably enhances its immunogenicity. Similar experiments suggest that macrophage-associated antigen enhances both primary and secondary antibody and cell-mediated immune responses: it is particularly effective in stimulating the production of B-memory cells and so primes the individual for a secondary response. Both the number of B lymphocytes responding to a weak antigen and the number of cells resulting from the clonal proliferation of each stimulated B cell are increased by macrophage participation. Macrophage activity does not, however, overcome the need for helper T cells in T-dependent antibody responses, nor does it induce immune responses in animals which have acquired immunological tolerance to the antigen.

There is recent evidence that macrophages secrete various non-antigenic products which influence lymphocyte responses to antigen, and also that macrophages can bind a product (? antigen receptors) of responding T lymphocytes (Fig. 5.27c, p. 130).

Trapping of lymphocytes by macrophage-associated antigen. When macrophages are treated *in vitro* with antigens and then washed free of unbound antigen

and added to a suspension of lymphocytes, a small proportion of the lymphocytes bind to individual macrophages to form clusters. The number of clusters is much greater when the lymphocytes are taken from an animal previously immunised with the same antigen, and it may thus be concluded that specifically-responsive lymphocytes, including memory cells, bind to the surface-bound antigen on the macrophage, from which they receive antigenic stimulation. The clusters contain T and B lymphocytes and it may be that, by attracting both types of cell, the macrophage promotes T-B cell co-operation in the immune response.

The distribution of antigen within the body depends on the route of entry and on the amount and nature of the antigen. When a small amount of antigen is introduced into the skin, it is taken up mainly by macrophages in the draining lymph nodes. Antigen entering the bloodstream is taken up by macrophages in the spleen and liver and also in various lymph nodes. These patterns of distribution are, however, influenced by whether the antigen is in solution or in particulate form, and by the presence or absence of antibodies resulting from a previous exposure to the antigen.

The virgin lymphocytes responsible for a primary immune response are believed to lie mainly in the spleen and lymph nodes (p. 122) where they presumably encounter the macrophage-bound antigen, although the possibility that these short-lived cells recirculate has not been ruled out. The secondary response is due to the presence, in the recirculating lymphocyte pool, of T and B memory cells responsive to the antigen (p. 123), and within a day or so of introduction of the antigen these specifically responsive memory cells aggregate at the sites of the macrophage-associated antigen, i.e. in the spleen and/or lymph nodes, and so disappear from the recirculating pool. Electron-microscopic studies have shown aggregates of lymphocytes around antigen-containing macrophages, similar to the *in vitro* clusters described above. In some instances, antigen persists locally in a non-lymphoid tissue, where its presence stimulates a macrophage reaction followed by aggregation of lymphocytes and a local immune response (p. 139).

Dendritic germinal-centre cells are not true macrophages, for they are of mesodermal origin. Their property of binding antigen-antibody complexes is described on p. 138.

Conclusions. *By destroying excess antigen and presenting the remainder in strongly immunogenic form, macrophages may inhibit the development of immunological tolerance and enhance the immune response. By their prolonged retention of antigen within the secondary lymphoid*

tissues, macrophages afford good opportunity for responsive T and B cells of the recirculating lymphocyte pool to encounter and respond to the antigenic stimulus. They thus help to ensure that

the immunological experience of the individual, represented by T and B memory cells, is brought to bear on any foreign antigen entering the body.

Histological Features of the Immune Response

The cellular events of immune responses take place mainly in the secondary lymphoid organs, i.e. the lymph nodes, Malpighian bodies of the spleen, tonsils and gut-associated lymphoid tissues. Study of the morphological changes of immune responses is not easy, for the large numbers of micro-organisms in the alimentary and upper respiratory tracts, and on the other exposed mucous membranes, provide continual antigenic stimuli which ensure that the immunity system is never completely inactive. The maintenance of animals in a germ-free environment from birth onwards is helpful, but technically exacting, and does not ensure complete freedom from antigenic stimulation: thus 'germ-free' mice are likely to be infected with mouse leukaemia virus, and do, in fact, show evidence of immune responses in their lymphoid tissues. There is also great regional variation in lymph node responses: when the antigenic stimulus is localised, e.g. in vaccination, the draining lymph nodes usually show the greatest response, but there may be marked differences, even between adjacent nodes.

Study of the lymphoid tissues in animals rendered deficient of T cells (e.g. by neonatal thymectomy, anti-lymphocyte serum or thoracic duct drainage), in fowls rendered B-cell deficient by bursectomy, and in patients with major congenital immunodeficiencies, has helped to distinguish between the histological features of the T-cell-mediated immune response and those of the B cell antibody response.

Lymph nodes

Normal structure. The lymph node consists of cortex, medulla and lymph sinuses (Fig. 5.30). The cortex occupies the superficial part of the node except at the hilar region: the medulla lies centrally, but extends to the hilum. The framework of the node consists of a net-

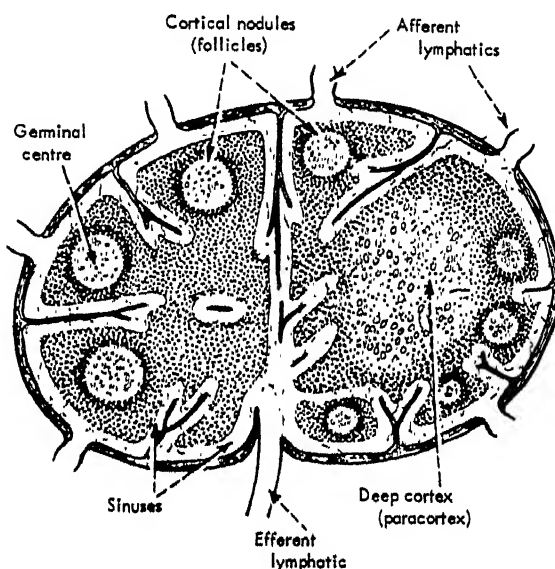


Fig. 5.30 Diagram of a lymph node, portraying the superficial cortex with nodules and germinal centres, and on the right an ill-defined area of deep cortex.

work of fine reticulin fibrils which are covered by the cytoplasm of elongated, flat reticulum cells with branching cytoplasmic processes. Many different names have been suggested for these cells, but their origin, nature and function are unknown. The sinuses are simply channels in the reticular framework, and they also are lined and traversed by reticulum cells. Macrophages are present in varying concentrations in all parts of the lymph nodes, including the lumen and walls of the sinuses.

Most of the free cells in the node are lymphocytes, and small lymphocytes usually predominate. In the cortex, the lymphocytes are closely arranged, and the superficial part of the cortex consists of foci, termed *primary nodules*, in which lymphocytes are more closely packed. In a stimulated node, as described below, a focus of lymphopoiesis, termed a *germinal centre*, may develop within the primary nodules. The deeper cortex, sometimes termed the

paracortex, consists of ill-defined uniform areas of cortical tissue lying between the superficial cortex and medulla (Fig. 5.31); in the stimulated node, there may be intense proliferation of lymphoid cells here, and this may result in one or more large cellular zones which compress the medulla of the node.

Lymph arriving at the node by the afferent lymphatics enters the peripheral sinus which surrounds the lymphoid tissue of the node and communicates at the hilum with the efferent lymphatic. From the peripheral sinus, cortical sinuses pass radially inwards to the medulla, running between the superficial cortical nodules and penetrating the deep cortex. In the medulla the lymph sinuses are numerous and the lymphoid tissue lies between them as the *medullary cords*: the medullary sinuses unite to form the efferent lymphatic.

The lymph nodes have two major functions. One of these is the interception and removal of abnormal or foreign material in the lymph stream passing through them, and is described on pp. 568–9, 573: the other is the production

of immune responses, the histological features of which are described below.

Immune responses in lymph nodes

(a) **Cell-mediated immune responses.** In lymph nodes draining the site of an antigenic stimulus of a kind which induces this type of response, e.g. an allogeneic skin graft or application to the skin of the hapten dinitrochlorobenzene, the most conspicuous early change is the appearance of large basophilic (and pyroninophilic) T '**blast**' cells or **immunoblasts** in the deep cortex of the node. These cells appear 2 days or so after antigenic stimulation; they multiply rapidly and are maximal at about 5 days. Small lymphocytes also increase in number in the deep cortex, which becomes enlarged and conspicuous, causing appreciable increase in size of the node. If the antigenic stimulation is not prolonged, the immunoblasts disappear after a further few days, the main feature then being an increased number of small lymphocytes in the deep cortex.

As explained on p. 121, the deep cortex is in the pathway of the recirculating T lymphocytes, which are responsible for cell-mediated immune responses, and the appearance of immunoblasts and subsequently small lymphocytes in the deep cortex represents the clonal proliferation of specifically responsive T cells following antigenic stimulation. The responsive cells may have encountered antigen in the tissues and passed to the lymph nodes, or antigenic material, either free or carried by macrophages, may have reached the draining nodes and there stimulated appropriately responsive cells among the recirculating lymphocytes passing through the deep cortex. The use of tritiated-thymidine labelling has shown that the T immunoblasts in the deep cortex proliferate to produce small T lymphocytes, and that these appear in the recirculating lymphocyte pool at about the time of development of cell-mediated immunity: recirculation will allow such 'primed' cells to encounter, and react with, the corresponding antigen almost anywhere in the body, a factor of obvious importance in combating infections.

(b) **Antibody production** involves the proliferation of B lymphocytes to provide B memory cells and the plasma cells which synthesise and secrete antibody. As with T cells, the recircula-



Fig. 5.31 Part of the cortex of a lymph node, showing two large cortical nodules (follicles) with germinal centres. The ill-defined area to the left of these consists of deep cortex. $\times 40$.

tion of B lymphocytes provides opportunity for their encounter with antigens in most of the tissues of the body. The main site of cell proliferation observed during antibody responses is in the superficial cortical nodules, which develop large spherical or ovoid **germinal centres** (Fig. 5.22), consisting of actively dividing, large basophilic B immunoblasts. **Plasma cells** appear in increasing numbers deep to the germinal centres and in the medullary cords, which are the main site of antibody production. During the early stages of the primary response neither antigen nor immunoglobulins are readily demonstrable in the cells of the germinal centres but, as antibody appears, antigen-antibody complexes become bound to the surface of the large **dendritic cells** of the germinal centres, and may persist there for many weeks. The nature of the dendritic cells, and the role of their surface complexes in antibody production, are not known; it is tempting to assume that the complexes provide a persistent antigenic stimulus to B cells, with consequent development of plasma cells which pass to the medulla and produce antibody, but this lacks proof. Small numbers of macrophages are also present in the germinal centres and contain ingested pyknotic nuclear material, the significance of which is unknown.

Although the function of the germinal centres is still uncertain, the intensity of mitotic activity within them during antibody responses suggests that they are the major site of clonal proliferation of specifically responsive B cells.

During antibody responses, which are often most pronounced in the lymph nodes draining the site containing antigen, the proliferating B cells differentiate into both plasma cells and memory B cells. Some of the B immunoblasts leave the nodes via the efferent lymphatic, and are distributed by the blood to other lymph nodes, etc., to the haemopoietic marrow, the lamina propria of the gut, and to the inflammatory reaction which may develop if antigen persists at the site of its introduction in the tissues (see below). Presumably these migrant cells also give rise to both plasma cells and B memory cells.

Although the histological features of cell-mediated immune responses and antibody production are described above separately, it must be emphasised that many antigenic stimuli induce both responses, and indeed the control

of antibody responses by helper and suppressor T cells (p. 129) requires the combined form of response, although the site of T-B cell interaction in the lymphoid tissues is not known.

Immune responses in other lymphoid tissues

Spleen. The structure of the spleen is described briefly on p. 561. Immune responses take place in the Malpighian bodies in which the area immediately adjacent to the central arteriole is occupied by T lymphocytes of the recirculating pool, and corresponds to the deep cortex of lymph nodes. The more peripheral lymphoid tissue is occupied by recirculating B lymphocytes: it corresponds to the superficial cortex of lymph nodes and is the site of formation of germinal centres during antibody production. Antibody-producing plasma cells appear at the periphery of the Malpighian bodies and pass into the adjacent red pulp.

Immune responses in the spleen occur particularly when antigenic material gains entrance to the blood stream, and present morphological appearances similar to those which have been described for the lymph nodes.

Gut-associated lymphoid tissues. The solitary lymphoid follicles and Peyer's patches of the gut and the lymphoid tissue of the appendix all have T- and B-cell areas, and are in the pathway of long-lived T and B lymphocytes (probably mostly or all memory cells—p. 123). The gut lymphoid tissue has no afferent lymphatics but lies immediately beneath the surface epithelium which in these sites is cuboidal and includes specialised 'M' cells which have a complex folded surface. Antigen in the gut lumen, including whole bacteria, can penetrate the overlying epithelium and stimulate an immune response in both T and B cells, with the usual histological changes, including germinal centre formation. The B cells of these lymphoid tissues do not specialise in the production of IgA antibodies, this being the function of plasma cells lying in the lamina propria of the gut mucosa and derived from other lymphoid tissues by way of the major lymphatics and blood (p. 123).

The thymus. As already indicated, the thymus is a site of T-cell lymphopoiesis. It is not an important site of immune responses, but occasional germinal centres and plasma cells

are demonstrable in the medulla of the thymus in a significant percentage of people dying suddenly or after a short illness: this suggests that B lymphocytes, presumably entering the thymus from the blood, are capable of mounting an antibody response. Thymic germinal centres are rarely seen in patients dying after a chronic illness, perhaps because there has been increased secretion of adrenal glucocorticoids, which results in thymic and lymphoid atrophy. Thymic germinal centres are, however, numerous in many cases of myasthenia gravis (p. 936), and occur also in the connective tissue diseases.

Immune responses in non-lymphoid tissues

In intensive and prolonged antibody responses, plasma cells may be widespread in various tissues, including the haemopoietic marrow, which can be an important site of antibody production. When antigen gains entrance to

non-lymphoid tissues, some of it is carried, either free or by macrophages, to the local lymph nodes where an immune response occurs. If, however, the antigen persists in its original site, as in chronic infections, skin allografts and locally-injected antigen of low solubility, then specifically primed T and B cells derived from the immune response in the lymph nodes may enter the tissue and respond specifically to the antigen: thus after 2 weeks or so there may be numerous plasma cells and also proliferating T lymphoblasts. Macrophages also accumulate and may play a role in the local immune response by processing the antigen (p. 134). If the antigen persists locally for some weeks, lymphoid tissue with germinal centres and T-cell areas may develop, i.e. ectopic secondary lymphoid tissue. This is seen also in the thyroid, etc. in the organ-specific autoimmune diseases (p. 162), and in the synovial membrane of joints in rheumatoid arthritis (Fig. 23.52, p. 918).

References

- Bach, J.-F., Dardenne, M., Pleau, J.-M. and Bach, A. A. (1975). Isolation, biochemical characteristics and biological activity of a circulating thymic hormone in the mouse and in the human. *Annals of the New York Academy of Science* **249**, 186–210.
- Burnet, F. M. (1959). *The Clonal Selection Theory of Acquired Immunity*. Cambridge University Press, Cambridge.
- Burnet, F. M. and Fenner, F. (1949). p. 76 in *The Production of Antibodies*, 2nd edn. pp. 142. Macmillan and Co. Ltd., London.
- Gershon, R. K. and Kondo, K. (1971). Infectious immunological tolerance. *Immunology* **21**, 903–14.
- Good, R. A., Martinez, C. and Gabrielsen, Ann E. (1964). Clinical considerations of the thymus in immunobiology. In *The Thymus in Immunobiology*, pp. 3–47. Ed. by R. A. Good and Ann E. Gabrielsen. Harper and Row, New York.
- Gowans, J. L. (1966). Life-span, recirculation and transformation of lymphocytes. *International Review of Experimental Pathology* **5**, 1–78.
- Jerne, N. K. (1955). The natural-selection theory of antibody production. *Proceedings of the National Academy of Science (N.Y.)* **41**, 849–57.
- Jerne, N. K. (1971). The somatic generation of immune recognition. *European Journal of Immunology* **1**, 1–9.
- Lawrence, H. S. (1969). Transfer factor. *Advances in Immunology* **11**, 196–266.
- Le Douarin, N. M. (1977). Ontogeny of primary lymphoid organs. In *B and T Cells in Immune Recognition*, pp. 1–19. Ed. by F. Looz and G. E. Roelants. John Wiley and Sons Ltd., Chichester.
- Miller, J. F. A. P. (1964). Effect of thymic ablation and replacement. In *The Thymus in Immunobiology*, pp. 436–60. Ed. by R. A. Good and Ann E. Gabrielsen. Harper and Row, New York.
- Mitchison, N. A. (1967). Immunological paralysis as a dosage phenomenon. In *Regulation of the Antibody Response*, pp. 54–63. Ed. by B. Cinader, pp. 400. C. C. Thomas, Springfield.
- Mitchison, N. A. (1971). The carrier effect in the secondary response to hapten-protein conjugates. *European Journal of Immunology* **1**, 10–27.
- Munro, A. and Waldmann, H. (1978). The major histocompatibility system and the immune response. *British Medical Bulletin*, **34**, 253–78.
- Owen, R. D. (1945). Immunogenetic consequences of vascular anastomoses between bovine twins. *Science* **102**, 400–1.

- Parrott, D. M. V. and de Sousa, M. A. B. (1971). Thymus-dependent and thymus-independent populations: origin, migratory patterns and lifespan. *Clinical and Experimental Immunology* **8**, 663–84.
- Williamson, A. R. (1979) Control of antibody formation: certain uncertainties. *Journal of Clinical Pathology*. Supplement (Royal College of Pathologists) **13**, 76–84.

Further Reading

- Advances in Immunology*. Vols. 1–28. (1961–1980). Academic Press, New York. (A continuing series of comprehensive articles on major immunological topics by leading authorities.)
- Bowry, T. R. (1977). *Immunology Simplified*, pp. 223. African Medical and Research Foundation. (A clear and brief account of basic and clinical immunology.)
- Herbert, W. J. and Wilkinson, P. C. (Eds.) (1977). *A Dictionary of Immunology*, 2nd edn., pp. 194. Blackwell Scientific, Oxford. (A most helpful compilation of brief descriptions of terms used in immunology.)
- Loor, F. and Roelants, G. E. (Eds.) (1977). *B and T Cells in Immune Recognition*, pp. 504. John Wiley and Sons, Chichester. (A collection of articles, by leading workers, on the development and functions of lymphocytes.)
- Roitt, I. M. (1977). *Essential Immunology*, 3rd edn., pp. 334. Blackwell Scientific, Oxford. (A clearly written and beautifully illustrated account of basic and clinical immunology.)
- Turk, John (Ed.) *Current Topics in Immunology Series*. Edward Arnold, London. (A series of monographs on basic and clinical aspects of important immunological topics.)
- See also bibliography for Chapter 6.

Immunopathology

This chapter is devoted entirely to disease processes which have an immunological basis. It falls naturally into two parts. First, the **hypersensitivity reactions** which are of an immunological nature. It should be noted that this use of the term hypersensitivity is somewhat restricted. It does not include those conditions in which the subject is abnormally sensitive to a drug as a result of genetically determined deficiency of an enzyme system necessary for metabolising the drug

or because of failure to excrete the drug or its metabolites, e.g. in diseases of the liver or kidneys: this type of undue responsiveness is termed *idiosyncrasy* and is not dealt with here.

The second main section of the chapter describes the **immunological deficiencies**, i.e. congenital or acquired conditions in which the subject is incapable of the normal range of immunological responses and as a result is unduly susceptible to infection.

Hypersensitivity Reactions

In most instances, hypersensitivity may be defined as a state in which the introduction of an antigen into the body elicits an unduly severe immunological reaction. It follows previous exposure to the antigen and is a consequence of the development of an immune response, i.e. production of antibodies or sensitised lymphocytes reactive with the antigen. *It is this reaction between the antigen and products of the immune response which produces the lesions of the hypersensitivity disease processes.*

Hypersensitivity reactions may be localised to the site of entry of the antigen, or generalised: the local reactions are mainly of an inflammatory nature, but may also include spasm of smooth muscle. The generalised effects include fever, shock, gastrointestinal and pulmonary disturbances, and sometimes fatal circulatory collapse. One of the earliest examples of hypersensitivity was provided by Richet and Partier (1902) who observed that intravenous injection of small amounts of extracts of sea anemone into dogs was harmless, but a second injection some weeks later was quickly followed by a violent and sometimes fatal reaction with

dyspnoea, vomiting, defaecation, micturition and collapse. Since this early report, which illustrates the acute and severe nature of some hypersensitivity reactions, a great deal has been learned, and hypersensitivity reactions may now be classified into four major types (see below).

The definition of hypersensitivity given above refers solely to **foreign antigens** entering the body from outside. However, the term includes also the conditions commonly known as the **autoimmune diseases**, in which antibodies or sensitised lymphocytes appear which are capable of reacting with a normal cell or tissue constituent *in vivo*, with consequent pathological changes. Hypersensitivity reactions may result also from passive immunisation, for example when antibody is produced in the mother by active immunisation by fetal red cells, and crosses the placenta in a subsequent pregnancy to gain entrance to the fetal circulation. Another special example of hypersensitivity of increasing importance is the **rejection process in allogeneic or heterogeneic tissue transplants**. The increasing diversity and use of

drugs has also provided an important group of **drug hypersensitivities** and the same applies to the expanding number of chemicals used domestically and in industry.

The four major types of hypersensitivity reactions are described briefly below and then each is dealt with in more detail. They have been elucidated very largely by animal experiments. Hypersensitivity reactions in man, whether they occur naturally or as a result of transplantation, or from administration of a drug, tend to be complex and often involve more than one of the four types.

Atopic, anaphylactic or type 1 reactions occur in individuals who are predisposed to develop increased amounts of IgE class antibodies in response to antigenic stimuli. IgE antibody binds to mast cells, and subsequent union of the corresponding antigen triggers off release of histamine, etc., from the sensitised mast cells, giving rise to a local inflammatory reaction and smooth muscle spasm, or to a more generalised reaction. Examples include hay fever and asthma.

Cytotoxic antibody or type 2 reactions occur when antibody develops which is capable of reacting with surface antigens of cells. As a result, the cells are injured by subsequent complement activation, phagocytosis, etc. Examples include destruction of red cells and platelets by auto-antibodies to their surface components.

Immune-complex, Arthus-type or type 3 reactions are caused by the reaction of antibody, usually of IgG class, with the corresponding soluble antigen. This can occur locally (Arthus reaction) or in the blood. In either case, immune complexes are deposited in the walls of blood vessels, where they activate complement and induce vascular injury.

Delayed hypersensitivity or type 4 reactions occur when the primed T lymphocytes, which develop during the cell-mediated immune response, encounter the corresponding antigen. The specifically reactive T lymphocytes transform to blast cells and secrete a number of factors (lymphokines) which mediate an acute inflammatory reaction, aggregation of more lymphocytes and monocytes, and sometimes necrosis. The tuberculin skin test is a good example.

It should be noted that type 1 and 2 reactions occur in subjects predisposed to unusual

immune responses, while types 3 and 4 are the result of immune responses of which all normal individuals are capable.

Before considering types of hypersensitivity in detail, this is a convenient place to give accounts of two systems, the profound importance of which is becoming increasingly apparent. The first is the **complement system**, which was for long regarded as being concerned solely in relation to antigen-antibody reactions, but is now known to participate in many other processes, e.g. in inflammation, blood clotting and fibrinolysis. The second is the **cyclic nucleotide system**, in which the levels of cyclic AMP and GMP within the cell play an important role in most of the functions of animal cells: its role in controlling the release of mast cell products, which is outlined briefly below, is involved in some forms of hypersensitivity, but is only one illustration of its much wider importance.

The complement system

This consists of at least 18 proteins which make up about 10 per cent of the total protein of the plasma. Many of the components are synthesised by monocytes and macrophages and increased macrophage activity, as in inflammation, is accompanied by enhanced production of some complement components.

Eleven of the complement components are termed C1–C9 (C1 is a complex of three factors, C1q, C1r and C1s), and when complement is activated by an antigen-antibody reaction these components react in sequential or cascade fashion in the order 1 to 9 except for C4 which is now known to react sequentially between C1 and C2. Such activation is termed the **classical pathway**. A second method of activation, termed the **alternative pathway**, is also initiated by antigen-antibody complexes and by certain bacteria without the necessity for an antigen-antibody reaction: it involves at least four additional factors—factors B, D, P (properdin) and C3b—and triggers off the sequential reaction at the C3 stage, so that C1, 4 and 2 are not involved. Various proteolytic enzymes present in inflammatory exudates and participating in the clotting and kinin systems can also activate the complement reaction at the C1 or C3 stages (p. 55). Complement activation is modulated by spontaneous decay of some of

the activated components and by a number of inhibitory factors, the best known of which are C1-INH which inhibits $\text{C1}\dagger$ and C3b-INA and $\beta 1\text{H}$, both of which inhibit C3b activity.

The classical pathway (Fig. 6.1) is initiated when IgG or IgM antibody reacts with antigen. As shown in Fig. 6.1, three of the stages of the pathway involve enzyme reactions, and since one molecule of enzyme can cleave many molecules of substrate, the sequential reaction is amplified as it progresses. Such amplification is, however, opposed by the rapid spontaneous decay of some of the activated components and by various inhibitory factors.

The alternative pathway. It is probable that there is continuous triggering of the classical pathway, perhaps by low concentrations of antigen-antibody complexes resulting from absorption of small amounts of environmental antigens or by low-grade plasmin activity. Such

complement activation is largely held in check by C1-INH , but presumably provides some C3b , which is an essential factor in the alternative pathway (Fig. 6.1). Normally, the alternative pathway is held in check by C3b-INA and $\beta 1\text{H}$, both of which are necessary to prevent uncontrolled cleavage of C3 by this pathway, and agents which trigger off the alternative pathway do so by interfering with the inhibitory function of C3b-INA and/or $\beta 1\text{H}$. The lipopolysaccharide cell wall material (including endotoxin) of various bacteria is the most important activator of the alternative pathway; it probably acts by binding C3b and rendering it resistant to the action of these inhibitors. The alternative pathway is also activated by IgA antibody-antigen complexes, and since C3b is the limiting factor in the alternative pathway, its production in the classical pathway also promotes alternative pathway activity.

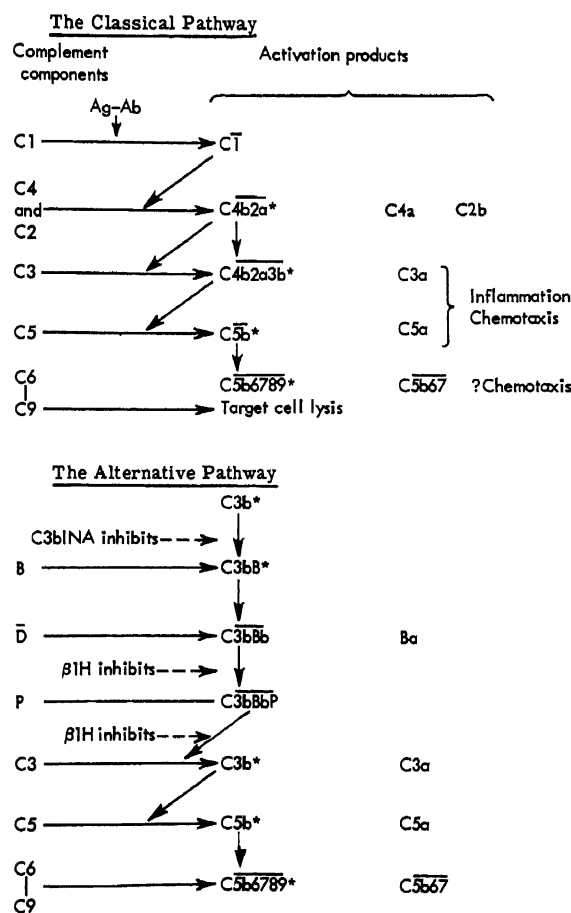


Fig. 6.1 Activation of complement. **The classical pathway** is initiated when C1 , which is a pro-enzyme, is activated by binding to the Fc of IgG or IgM antibody complexed with antigen (Ag-Ab). The active enzyme C1 (activated components of the system are indicated by overlining) cleaves both C4 and C2 into small fragments (C4a and C2b) and large fragments (C4b and C2a). The large fragments unite to form a second enzyme C4b2a which cleaves C3 into C3a and C3b . The large fragment, C3b , binds to C4b2a to form a third enzyme, C4b2a3b , which cleaves C5 into C5a and C5b . C6 and C7 then bind to C5b , forming a stable complex, C5b67 . C8 binds to this complex which in turn binds C9 to form the C5b-9 lytic complex.

The alternative pathway requires C3b for its initiation: this is provided by continuous spontaneous low-grade activity of the classical pathway, or by splitting of C3 by plasmin, the kinin or clotting systems, or by various proteolytic enzymes released by polymorphs, injured tissue cells, etc. C3b is inactivated by C3b-INA and $\beta 1\text{H}$, but is protected by the binding of B (and probably also by endotoxin). D splits B in this complex, releasing Ba , leaving C3bBb which binds P (properdin) to form a complex which splits C3 , producing more C3b : this amplifies the alternative pathway activation and also contributes to an enzyme complex which splits C5 to provide C5b . The reaction then proceeds as in the later stages of the classical pathway.

* Those activated components and complexes which bind to cell surfaces are marked with an asterisk.

† Activated components of the complement system are indicated by overlining, e.g. $\text{C1}\dagger$.

Biologically active products of complement. When complement is activated by either pathway, some of the activation products are released into the surrounding fluid. Of these, C3a and C5a influence the behaviour of various cells. They stimulate mast cells and basophils to release histamine and other vasoactive amines and thus induce vascular exudative changes as in acute inflammation. In these effects, C5a is far more potent than C3a. C5a is also chemotactic for neutrophil polymorphs and monocytes and so promotes emigration and accumulation of these cells. C3a and C5a are commonly, though inappropriately, termed *anaphylatoxins* (p. 55). When complement is activated by particulate material, e.g. micro-organisms or host cells sensitised with antibody, or certain bacteria alone, some of the activation products, indicated in Fig. 6.1 by an asterisk, bind to the surface of the target cell and promote its destruction by two methods. Firstly, polymorphs and macrophages have surface receptors for C3b, adherence of which to the target cell (*immune adherence*) thus promotes the binding of polymorphs and macrophages and favours phagocytosis and destruction of the target cell. Secondly, completion of the complement cascade results in the insertion of the C5b-9 complex into the target cell plasma membrane which is consequently injured, sometimes causing cell death. The mode of injury is not fully understood, but in electron micrographs apparent holes develop in the cell membrane (Fig. 2.10, p. 18), the cell absorbs water and electrolytes, swells up and ruptures.

When complement is activated by either pathway, C5b is released and may adhere to adjacent cells not involved in the initial activation. The C5b-9 complex may then build up on these 'innocent' cells, causing their death. This is known as **bystander cell lysis** or **reactive lysis**.

The pathological importance of complement. As explained earlier (pp. 44–61), complement is a source of *mediators of the acute inflammatory reaction* and appears to be of importance in the inflammation induced by various non-immunologic stimuli. The role of complement in types 2 and 3 hypersensitivity reactions is discussed later in this chapter.

Probably the most important role of complement is in killing micro-organisms and rendering them susceptible to ingestion and killing by phagocytes (pp. 180–2).

Complement deficiency. Genetically-determined deficiency of one or other of the complement components is rare. Deficiency of an early component (C1, C4 or C2) does not usually give rise to disease, presumably because activation can still occur by the alternative pathway. Deficiency of a later component usually results in an increased susceptibility to bacterial infection.

The best known genetic abnormality of complement is deficiency of C1-INH which results in uncontrolled activation of complement and the inflammatory lesions of hereditary angio-oedema (p. 254).

Detection of complement components. The measurement of total haemolytic complement activity present in serum is performed traditionally by determining the concentration of serum required to cause lysis of 50% of a suspension of red cells sensitised with antibody (p. 112). Techniques are, however, available to measure individual complement components, and also their activation products: such tests are now in routine use to detect evidence of activation of complement *in vivo* in various diseases, and to distinguish between activation by the classical and alternative pathways.

The cyclic nucleotides

The integrity of multicellular organisms is obviously dependent on co-ordinated function of their individual cells. This involves systems of communication between cells by the nervous system, and the responsiveness of cells to the levels of hormones and various other solutes in the extracellular fluid. A major advance in the understanding of how cells respond to such signals has been provided by the discovery that many cell functions are controlled by the concentration of cyclic nucleotides in the cytosol (see Larner, 1977). The effects of these compounds are illustrated here by the role they play in release of stored products by mast cells, a phenomenon of importance in type 1 hypersensitivity.

The release of histamine and other vasoactive agents stored in mast cell granules has already been discussed in relation to the increased vascular permeability of acute inflammation. It is triggered off by various physical and chemical agents. In type 1 hypersensitivity, degranulation of mast cells (and basophil leukocytes) occurs when molecules of antigen

combine with molecules of IgE class antibody bound to the mast cell surface (see below). β -adrenergic receptor blockade also has the same effect, while β -adrenergic stimulation, e.g. by isoprenaline, salbutamol and adrenaline, inhibits mast cell degranulation (Fig. 6.2). Stimulation or blockade of α receptors and cholinergic receptors also influence mast cell degranulation. It is thus apparent that many agents have an effect on this important function of mast cells. It now appears that they operate through a common pathway—by raising or lowering the levels of cyclic nucleotides in the cytosol of the mast cell.

Agents which stimulate β receptors activate an enzyme, adenylyl cyclase, which lies at the inner surface of the cell membrane, and this results in increased production of cyclic adenosine monophosphate (cAMP) which suppresses release of stored products of the mast cell, while β -adrenergic blockade inhibits adenylyl cyclase with consequent fall in cAMP level and the stored products are released. Stimulation or blockade of the α -adrenergic receptors have the opposite effects. Agents which react with the cholinergic receptors influence the activity of guanylyl cyclase which in turn affects the level

in the cytosol of a second nucleotide, cyclic guanosine monophosphate (cGMP). In general, cAMP and cGMP have opposite effects on cell function, and it appears to be the ratio of the two which is of importance. Thus a fall in cAMP and/or a rise in cGMP favour degranulation. The cyclic nucleotides are inactivated by intracellular phosphodiesterases, and many agents, including some of those mentioned above, also influence the levels of the cyclic nucleotides by affecting the activity of these enzymes.

Not only do cAMP and cGMP have opposite effects on the cell, but some agents which lower the level of one increase the other. This may be mediated by an effect on the permeability of the cell membrane for Ca^{++} or on the distribution of Ca^{++} within the cell, for a rise of Ca^{++} in the cytosol inhibits adenylyl cyclase activity and activates cAMP-phosphodiesterase, thus decreasing cAMP: it has the opposite effect on cGMP.

Although the mast cell has been used to illustrate the importance of the cyclic nucleotides, they are of equal importance in all cells, exerting an influence on cell division, motility, secretion, etc. A fall in cAMP and rise in cGMP is

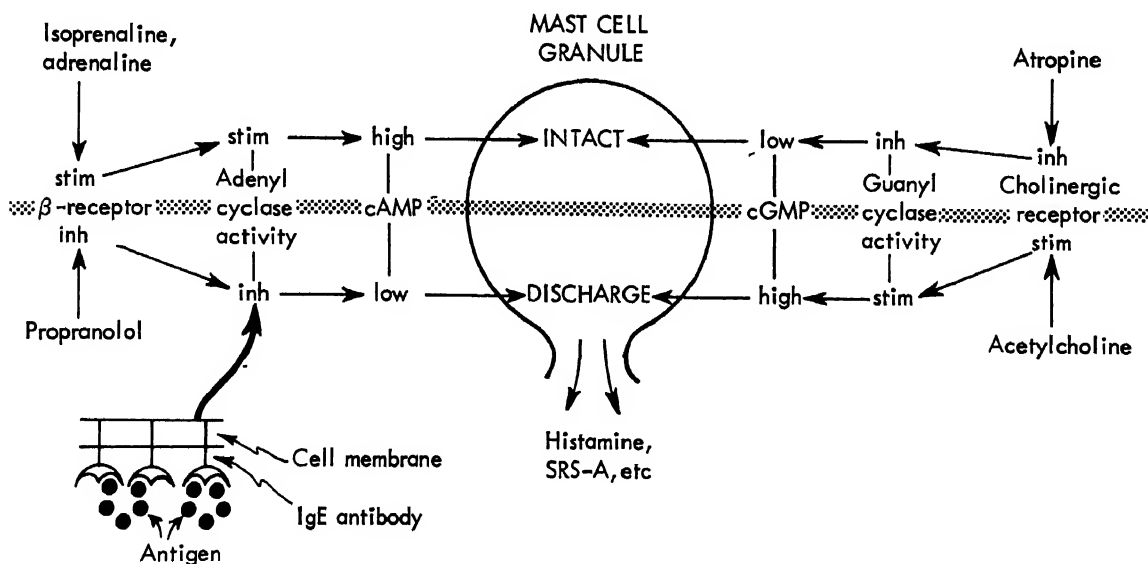


Fig. 6.2 Discharge of mast-cell and basophil-leukocyte granules, showing the relationship between the levels of the cyclic nucleotides (cAMP and cGMP) in the cytosol and release of products stored in the cell granules. Stimulation (stim) and inhibition (inh) of β -adrenergic and cholinergic cell surface receptors influence the activity of adenylyl and guanylyl cyclases and thus the levels of cAMP and cGMP in the cytosol. The attachment of antigen to cell-bound IgE antibody inhibits adenylyl cyclase and causes discharge of the granules. Although α -adrenergic receptors (not shown) also influence the levels of cyclic nucleotides, the mechanism has not been elucidated. Note that the pathways above the hatched lines stabilise mast cell granules and those below stimulate their discharge.

associated with mitosis. In many glandular cells, the same change is associated also with secretory activity, although in some instances, e.g. the cells of the adenohypophysis and the thyroid epithelium, secretion is induced by a rise in cAMP and a fall in cGMP.

Atopy (anaphylactic, 'immediate' or type 1) hypersensitivity

Approximately 10% of the population suffers from this type of hypersensitivity, although in most of these the symptoms are mild and occasional. The commonest manifestations of atopy are **hay fever** and **extrinsic asthma**, which tend to run in families and are sometimes preceded by **atopic eczema** in infancy and childhood. The sufferer from hay fever develops acute inflammation of the nasal and conjunctival mucous membrane with sneezing and nasal and lacrimal hypersecretion within minutes of exposure to an atmosphere containing the causal agent (usually grass or other pollens). Similarly, an acute attack of asthma, with difficult wheezing respiration due to narrowing of the airways by bronchospasm and mucous secretion, develops rapidly when the asthmatic inhales the agent to which he is hypersensitive, e.g. house dust or animal dander. Atopic individuals, particularly in childhood, may also suffer from '**food allergies**' in which absorption of antigenic constituents of certain foods, e.g. milk or eggs, promotes an acute reaction in the gut with colicky pain, vomiting and diarrhoea.

Urticaria, consisting of acute inflammatory lesions of the skin with wealing due to dermal oedema, is common in atopic subjects and also occurs alone as an acute or chronic condition.

In addition to local disturbances, atopic patients sometimes develop **acute systemic anaphylaxis** (anaphylactic shock) with dyspnoea, urticaria, convulsions, prostration and sometimes death. Generalised reactions occur when the responsible agent is absorbed in amounts which produce a significant level in the blood. Fortunately, severe anaphylactic shock is rare, but it sometimes occurs in hypersensitivity to drugs, notably penicillin, and to the venoms of stinging insects.

Skin tests and provocation tests. Diagnosis of atopy depends firstly on an accurate clinical history, which usually suggests that acute attacks result from exposure to a particular

environmental antigen. To confirm the state of hypersensitivity, dilute solutions of the suspected antigens (which are commercially available) may be placed on the skin and pricked in with a needle. A positive result is indicated by a local weal and flare reaction, developing within a few minutes (Fig. 6.3) and lasting for an hour or so. Hence the term 'immediate type' hypersensitivity. Although of diagnostic help, skin tests are not infallible. The atopic individual tends to give positive reactions not only to the environmental antigen(s) responsible for attacks of atopy, but also to various others. Also, in a proportion of cases the skin test is negative although the history is very suggestive of atopy to that antigen. A more reliable indication of the causal role of a particular antigen is provided by provocation tests, for example bronchial challenge by controlled inhalation of the suspected antigen by the asthmatic, and nasal application for the hay-fever patient: an acute attack implicates the test antigen. In all such tests, careful precautions must be taken to avoid provoking a severe reaction.

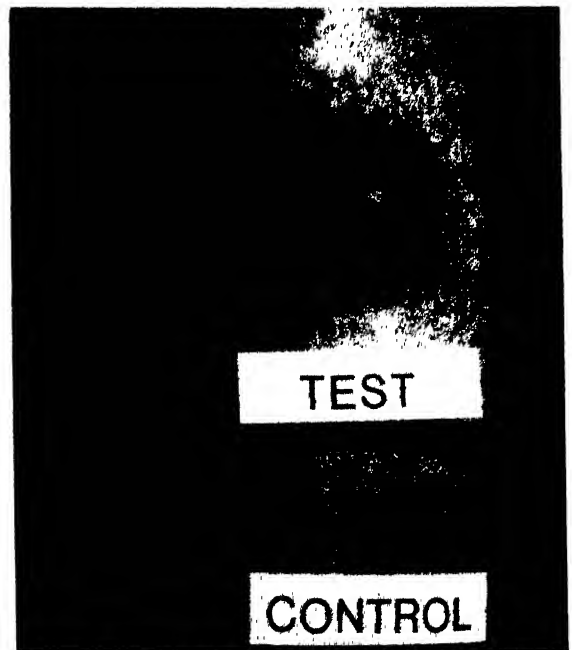


Fig. 6.3 Skin test showing immediate (type I) hypersensitivity reaction. The patient was an asthmatic and the test was performed by intradermal injection of an extract of house dust. Note the oedema and wide zone of reddening. Photograph taken 10 minutes after injection. (Dr. Ian McKay.)

The mechanism of atopic hypersensitivity

Passive transfer of immediate type hypersensitivity was demonstrated beautifully 60 years ago by the classical investigation of the two German doctors, Prausnitz and Küstner. Küstner himself regularly developed hypersensitivity reactions immediately after eating fish, and intradermal injection of an extract of cooked fish induced an immediate-type reaction. Injection of a small amount of Küstner's serum intradermally into Prausnitz induced a local state of hypersensitivity, for when cooked fish muscle extract was injected intradermally 24 hours later at the same site an immediate-type reaction occurred, whereas other skin sites were negative.

Passive transfer has been confirmed repeatedly with serum from atopic subjects: it has been shown to be antigen-specific, and the interval between the two injections can be prolonged to 3–4 weeks, demonstrating that the serum factor (reagin—see below) binds to some tissue element in the skin.

Temporary atopic hypersensitivity has also been observed in recipients of blood from atopic donors: the recipient exhibits positive skin tests and may develop clinical atopy if exposed to the relevant antigen(s). Evidence that reagin sticks to various tissues and not just to skin has been provided by the demonstration that fresh bronchial tissue removed from an atopic subject undergoes contraction of the smooth muscle when exposed to the appropriate antigen, and it has been shown that normal tissues of various types can be sensitised passively by the serum of atopic individuals.

Reaginic antibody: IgE. Being antigen-specific, the serum factor responsible for passive transfer of atopy has long been regarded as an antibody, and known as **reagin** or **reaginic antibody**. It has proved difficult to characterise, for it is present in serum in only trace amounts, and is relatively instable; also it is **homocytotropic**, *i.e.* binds to the tissues of man or related primates, but not to tissues of other genera: this restricts its detection by passive transfer to experiments on man and some monkeys. However, it was eventually shown by Ishizaka *et al.* (1966) to be due to an immunoglobulin of a distinct 'new' class, since termed IgE. Patients with IgE-producing mye-

lomas (plasma-cell tumours) have provided a rich source of IgE. Using this material, it was shown that when a solution of IgE or of its Fc component was injected into the skin it was found to block the tissue sites of attachment of reaginic antibody, and so inhibited the Prausnitz-Küstner reaction at the same site. This is strong confirmation that reaginic antibodies are of IgE class, and shows that fixation to tissues is a property of their Fc component. Antibody specific for myeloma IgE is now used to assay the level of IgE in serum and also as the basis of the *in-vitro* assay of specific IgE antibodies, *e.g.* by the 'radio-allergosorbent test' (RAST). Raised levels of IgE, and of IgE class antibodies to the relevant antigens, have been detected in the serum and nasal secretions, etc., but such *in-vitro* tests have so far contributed little to clinical practice.

Role of mast cells and basophil leukocytes. Mast cells are widely distributed in most tissues, and are particularly numerous adjacent to small blood vessels. Basophil leukocytes resemble mast cells in appearance and function; like other leukocytes, they can respond to chemotaxins and migrate from the blood into the tissues. Both mast cells and basophils have large basophilic cytoplasmic granules which can be discharged by various stimuli, *e.g.* the various causal agents of acute inflammation, and which release histamine and other vaso-active compounds.

Many mammalian species are capable of developing immediate-type hypersensitivity reactions similar to atopic reactions in man, and animal studies have demonstrated that reaginic antibody binds firmly by its Fc component to surface receptors of mast cells and basophils. This leaves the Fab ends of the antibody free to react with the corresponding antigen, and when this occurs, cross-linking of surface-bound antibody molecules by the antigen results in degranulation via the cyclic-nucleotide pathways (p. 145), and the discharged granules in turn release their stored histamine and other vaso-active agents which bring about vascular hyperaemia and exudation (Fig. 6.4). Prior depletion of mast cell and basophil mediators by such compounds as 48/80 inhibits the immediate hypersensitivity reaction in animals, as does the administration of histamine antagonists.

Although atopy in man differs in certain

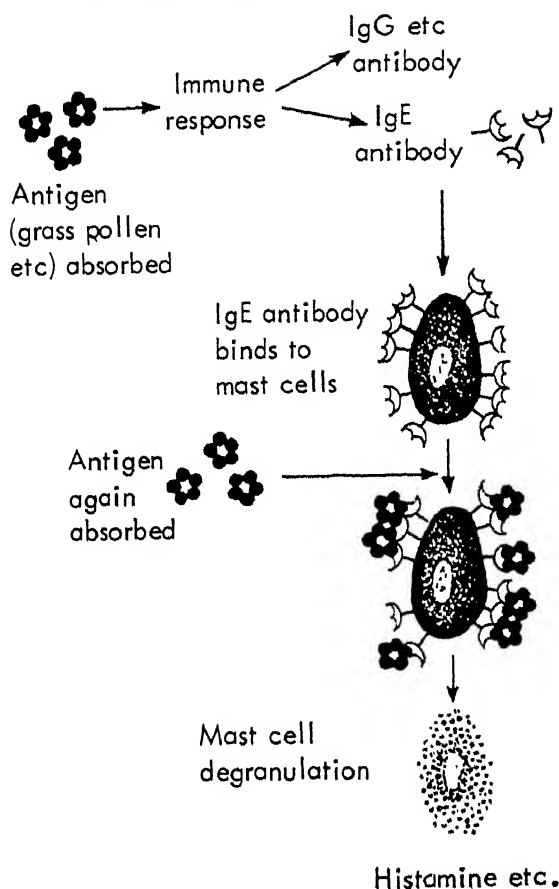


Fig. 6.4 The mechanism of atopic reactions. IgE antibody to pollens, etc., binds by its Fc component to mast cells (and basophil leukocytes), and subsequently absorbed antigen triggers off the sensitised mast cells (probably by linking together antibody molecules on their surface) to release histamine, etc.

respects from experimental immediate-type hypersensitivity in animals, and has not been so thoroughly studied at cellular level, there is nevertheless strong evidence for the mechanism summarised in Fig. 6.4. Human basophil leukocytes have been shown to bind IgE antibodies and to be degranulated by subsequent addition of antigens. Mast cells in monkey lung have been shown to bind human IgE, subsequent cross-linking of which causes the cells to release histamines, etc. (Ishizaka, Ishizaka and Tomioka, 1972). I am unaware of similar work on human mast cells, but it has been shown that human lung tissue removed surgically from an atopic subject (who had developed bronchial carcinoma) yielded histamine and SRS-A (the slow reacting substance of anaphylaxis) when perfused with fluid containing a

solution of the appropriate antigen. The main source of these two vaso-active agents in man are the mast cells and basophils. SRS-A is a lipid which increases vascular permeability and induces more prolonged spasm of bronchial smooth muscle than does histamine. Other factors released by mast-cells and basophils include two tetrapeptides known collectively as ECF-A (the eosinophil chemotactic factor of anaphylaxis) and possibly prostaglandins.

The role of eosinophil leukocytes. Acute atopic reactions are characterised by vascular hyperaemia, exudation, and emigration of eosinophil polymorphs into the affected mucous membrane. The role of eosinophils has long remained a mystery, but evidence is accumulating that they modulate the intensity of atopic reactions. Experimentally-induced immediate-type hypersensitivity reactions in rats are enhanced by injection of an anti-eosinophil serum which depletes the blood of eosinophils, and these cells not only secrete histaminase and arylsulphatase, which inactivate histamine and SRS-A respectively, but there is also evidence that they may inhibit both mast-cell degranulation and restoration of degranulated mast cells. The eosinophil also plays a defensive role against parasitic worms (see below).

Other immunological factors. Although IgE antibody is the important mediating agent of atopy, it may not be the only such factor. There is recent evidence that IgG4 antibody (p. 108) can mediate a complement-dependent form of immediate type hypersensitivity, although it has not yet been proved that this second mechanism plays a significant role in human atopy. The antibody response is not restricted to IgE: other classes of antibody are also produced and in some atopic subjects the immediate reaction is followed by an Arthus (type 3) hypersensitivity reaction due to the formation of antigen-antibody complexes and possibly also by a delayed-hypersensitivity (type 4) reaction, due to primed T lymphocytes, in some cases. These later reactions, which may cause difficulty in detecting the responsible antigen(s), have been demonstrated most readily by provocation tests.

Atopic hypersensitivity can often be reduced, but seldom abolished, by subcutaneous injection of the responsible antigen in a slowly soluble (e.g. alum precipitated) form. The mechanism of such **hyposensitisation** is not known,

but it has been shown to lead to a fall in IgE antibody and a rise in IgG antibody in the serum over the next few months. It is thought that the IgG antibody has a 'blocking' effect by reacting with naturally-absorbed antigen, thus preventing its combination with IgE antibody bound to mast cells. Hyposensitisation is not always successful, and the outcome is not predictable. There is, moreover, a possibility that, by stimulating the production of IgG antibody, the procedure may predispose the individual to immune complex disease (p. 152). The possible role of IgA antibodies in atopy is discussed below.

Predisposition to atopy

The occurrence of atopy in several generations of some families suggests a genetic predisposition. In general, atopic subjects have higher total serum IgE levels, and produce more IgE antibody in response to antigenic stimulation, than control subjects. Accordingly it seems likely that the genetic factor determines the intensity of IgE responses. Some support for this is provided by the finding of an association, in some families predisposed to atopy, between the IgE response to ragweed antigen (commonly a cause of hay fever) and the possession of particular HLA antigens.

Although atopy is attributable to the reaction between antigens and IgE antibodies, there must be other important factors, for in some cases the skin tests do not correlate with provocation tests nor with the occurrence of clinical atopy. One possibility is that the affected mucous membranes, e.g. the nasal mucosa in hay fever subjects and the bronchial mucosa in asthmatics, are unduly permeable to small antigenic molecules (most of those which induce atopy have a molecular weight of 40 000 or less), but the evidence for this is not convincing. Related to this is the possible importance of the IgA antibody response and the secretion of IgA by mucous membranes. It has been shown that injection of IgA antibody into rats inhibits the absorption of inhaled or injected antigens. Most atopic subjects have normal serum IgA levels, but it has been reported that intranasal application of antigen results in the appearance of less IgA antibody and more IgE antibody in the nasal secretion in hay-fever subjects than in controls (Butcher, Salvaggio

and Leslie, 1975). It is also of great interest that assay of the serum IgA levels of the infants of atopic parents has shown that those with a low level of IgA at three months of age are more liable subsequently to develop atopic eczema (Taylor *et al.*, 1973) and probably also asthma (Soothill, personal communication). The same group of workers has also reported that avoidance of the environmental antigens which commonly cause atopy (including dairy products) during the first six months of life reduces the incidence of atopic eczema in infants with an atopic predisposition (Matthew *et al.*, 1977).

It has also been reported that T-cell depletion in rats enhances IgE antibody responses. In atopic subjects, there is no good evidence of a T-cell deficiency, although children with the rare Wiskott-Aldrich syndrome (p. 171), who have a congenital T-cell deficiency, are prone to develop atopy.

There is also evidence that parasympathetic bronchoconstrictor nerve endings between the bronchial epithelial cells are unduly irritable in asthmatics. These nerve endings are triggered by various stimuli, possibly including antigen-antibody complexes, and it has been postulated that an asthmatic attack may be initiated by formation of such complexes with consequent stimulation of cholinergic mast cell receptors and degranulation (Fig. 6.2). From these various observations, it is apparent that a number of abnormalities in the pathway between the exposure of the mucosa to antigens and the reaction of submucosal blood vessels, mucous glands and smooth muscle of the bronchi could be involved in atopy, but none have been demonstrated with certainty to be of importance.

It is noteworthy that in African communities with a high rate of infestation with parasitic worms, IgE levels are also high, although atopy is uncommon. Infestation with worms or injection of worm extracts has been shown in animals to increase the IgE antibody response to various antigens. IgE antibody is important in the defence against worms, and it has been suggested that, by binding to mast cells, anti-parasitic IgE antibody, if present in relatively high concentration, excludes the binding of other IgE antibodies and thus protects against atopy. This would account for the higher incidence of atopy in populations with a low level of parasitic infestation, and it may be that elimination of most of our parasites has exposed us

to the harmful effects of IgE antibody responses to various otherwise harmless environmental antigens.

Cytotoxic antibody (type 2) reactions

The only distinctive feature of this type of hypersensitivity reaction is that it is mediated by antibodies which cause injury to cells by combining specifically with antigenic determinants on their surface. With few exceptions, the targets of cytotoxic antibodies are the cells of the blood. Such injury has been investigated mainly in man, in whom it may occur in the following circumstances.

1. *Auto-antibodies* may develop which are reactive with normal antigenic constituents on the surface of cells. This unexplained breakdown of self-tolerance may occur in isolation or as a feature of systemic lupus erythematosus (p. 164) and is sometimes associated with a number of infections and with lymphocyte neoplasia.

2. *Drug-induced cytotoxic antibodies.* Some drugs or their metabolites bind to the surface of one or other type of cell: if such a drug is haptenic, it induces an antibody response, and the reaction of antibody with the cell-bound hapten may bring about destruction of the cell.

3. *Iso-antibodies* can cause injury to cells of the blood following blood transfusion or transplantation of haemopoietic or lymphoid tissue. Maternal iso-antibodies of IgG class may also pass through the placenta and injure the cells of the fetus.

Cytotoxic antibodies to cells of the blood

Cytotoxic auto-antibodies. The classical example is *auto-immune haemolytic anaemia* in which red cell injury is brought about by auto-antibody reactive with various antigenic determinants inherent in the surface of red cells. The antibody may be of IgG or IgM class, and can be detected on the red cell surface by the antiglobulin test (p. 111). When present in low concentration on the cell surface, IgG may have little or no effect. In higher concentration, it promotes the binding of the red cell to macrophages which have receptors for the Fc of IgG; such binding may result in injury to the red cell membrane or to phagocytosis and de-

struction of the red cell by macrophages, mostly in the red pulp of the spleen and the hepatic sinusoids. IgG antibody can also activate complement and cause intravascular lysis of the cells: this requires pairs of IgG antibody molecules bound to closely adjacent antigenic sites on the red cell surface. IgM antibodies, even in low concentration, often cause red cell destruction by intravascular lysis: single IgM molecules binding to two or more antigenic determinant sites on the red cell surface are capable of activating complement. Complement activation also promotes binding of the red cells to macrophages (by the C3b receptors of the latter), while IgM antibody can also cause agglutination of red cells, particularly where the circulation is slow, as in the red pulp of the spleen; both these effects result in intra-splenic destruction of red cells.

Idiopathic thrombocytopenic purpura is caused by auto-antibody which reacts with the surface components of normal platelets, with similarly destructive effects. The frequency with which splenectomy is followed by a rapid rise in the platelet count indicates the importance of the splenic macrophages in the increased platelet destruction.

Auto-antibodies to leukocytes may be a cause of leukopenia (reduced numbers of leukocytes) but this is difficult to prove, partly because such auto-antibodies must usually be sought by testing the patient's serum with leukocytes from another individual, and iso-antibodies to leukocytes are a common snag. Secondly, leukocytes tend to bind IgG non-specifically and give a false-positive antiglobulin test.

Drug-induced cytotoxic antibodies. Some drugs or their metabolites are capable of binding to the surface of red cells, leukocytes or platelets and acting as haptens. Antibody develops and binds to the hapten on the cell surface, and cell injury and destruction may then result, as in the case of auto-antibodies (see above). A good example of this is provided by penicillin, the benzyl-penicilloyl degradation product of which binds firmly to red cells. Most people who have received penicillin have some antibody (usually IgM) to the penicilloyl group, but after prolonged heavy dosage, high titres of IgG antibody develop in some patients and this brings about the destruction (mostly by splenic phagocytosis) of sensitised red cells. Some drugs, for example rifampicin, result in

platelet destruction by an immunological reaction; this may be due to its binding to the platelet surface and acting as a hapten, as described above, but it seems more likely that it forms complexes with antibody in the plasma and that it is the binding of such complexes to Fc receptors of the platelets which causes their destruction (p. 160).

A very few drugs can induce the development of *auto-antibodies*: for example, patients receiving α -methyl dopa for a few months often develop auto-antibodies to surface antigens on their red cells, detectable by the direct antiglobulin test: there is no evidence that the antibodies react with drug-derived antigens, for they are true auto-antibodies and react with the patient's and other individuals' red cells. In most instances, there is insufficient antibody to cause significant red cell destruction, but approximately 1 per cent of patients develop a haemolytic anaemia which gradually disappears on withdrawing the drug; the mechanism of auto-immunisation is obscure.

Cytotoxic iso-antibodies. In blood transfusion, administration of red cells possessing the A or B surface iso-antigens to an individual whose plasma contains the natural anti-A or anti-B iso-antibodies usually results in rapid destruction of the donated red cells. These natural antibodies are of IgM class and so complement fixation and lysis of the incompatible red cells results. There are literally dozens of other red cell iso-antigens, but normally the corresponding antibodies appear in the plasma only after blood transfusion or pregnancy (see below). After the ABO groups, the Rhesus (Rh) system of iso-antigens is of most importance in man. Transfusion of red cells possessing Rh antigens which are not present in the recipient's red cells often results in development of the corresponding Rh antibody, following which the transfused cells are destroyed abnormally rapidly in the spleen. During labour (or abortion) some fetal red cells enter the mother's circulation, and, if Rh-incompatible (which depends on the father's Rh group) they sometimes stimulate development of Rh antibodies.

The Rh antibodies are particularly important in pregnancy, because they are usually mainly of IgG class, and so can cross the placenta. If the pregnant woman has developed Rh antibodies, as a result of a previous pregnancy or

blood transfusion, they enter the fetal circulation and, if the fetal red cells possess the corresponding Rh antigens, abnormal destruction results in fetal death or anaemia (Fig. 6.5). Once an individual has developed Rh antibodies, from either pregnancy or transfusion, subsequently transfused Rh-positive red cells are liable to be destroyed rapidly. The development of Rh antibodies can very often be prevented by injecting Rh antibody into the Rh-negative woman within 48 hours after termination of an Rh-incompatible pregnancy.

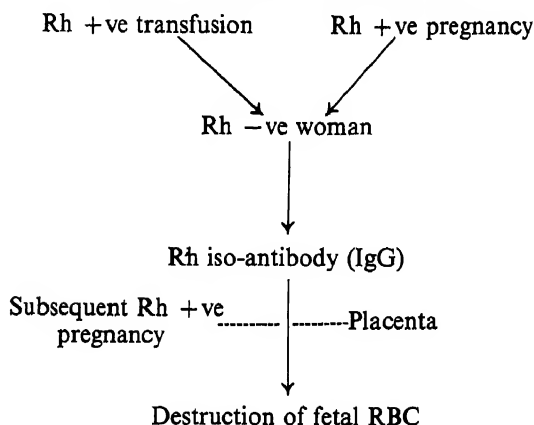


Fig. 6.5 Haemolytic disease of the newborn. Fetal red cell destruction is brought about by maternal isoantibody which has developed as a result of previous Rh +ve pregnancy or transfusion of Rh +ve blood.

In common with tissue cells, the leukocytes and platelets have antigens belonging to the HLA and other 'transplant antigen' systems (p. 166). This is seldom of importance in blood transfusion unless the aim is to supply these cell types to a deficient recipient. Blood transfusion and pregnancy do, however, stimulate the development of HLA and other antibodies, and subsequently transferred platelets or leukocytes may be destroyed rapidly. Such iso-immunisation is of importance if subsequent transplantation, e.g. of a kidney, is performed (p. 846), and maternal iso-antibodies to platelets may also cause thrombocytopenia in the fetus.

Auto-antibodies to tissue constituents

Auto-antibodies which react with components of tissue cells *in vitro* are demonstrable in the serum of patients with various diseases (p. 161), but the available evidence suggests that they are not usually a major cause of cell injury. In

some instances, this is because they react with intracellular, as opposed to surface constituents, of the target cells. A good example is antibody to deoxyribonucleoprotein, which is present in the plasma of patients with systemic lupus erythematosus: it can react with, and lead to destruction of, the nuclei of dead cells, but does not reach the nuclei of living cells.

There are, however, two examples of auto-antibodies which react with cell surface receptors and have a profound effect on the target tissue cells. One is a thyroid auto-antibody, which reacts with the TSH receptor of thyroid epithelium and, like TSH itself, stimulates the cell to increased function and proliferation: this is the cause of *Graves' disease*, the common type of hyperthyroidism. The other is an antibody to the acetylcholine receptor in skeletal muscle cells: it blocks the receptor and thus causes the muscle weakness of *myasthenia gravis* (p. 936). A third example of a harmful auto-antibody is found in a small proportion of sterile men: it reacts with spermatozoa and may be present in sufficient concentration in seminal fluid to impair their motility.

Although 'cytotoxic' implies injury to cells, type 2 hypersensitivity is sometimes extended to include antibody-induced injury to extracellular tissue elements. The best known example of this is the rare type of glomerulonephritis in which auto-antibody develops to glomerular capillary basement membrane. Union of this antibody with the inner surface of the basement membrane is followed by activation of complement, as in the Arthus reaction, and a destructive inflammatory lesion results in the glomeruli (p. 832).

Antibody-dependent lymphocyte cytotoxicity

When a suspension of living cells is treated with an IgG class antibody which reacts with their surface membrane, and normal lymphocytes (e.g. from the peripheral blood of a normal individual) are added, some of the lymphocytes bind to the surface of the sensitised cells and bring about their destruction: this probably involves penetration of the target cell membrane by the lymphocyte (Reid *et al.*, 1979). The cytotoxic lymphocytes, sometimes called K cells (p. 120), do not have surface Ig, and so are not B lymphocytes: they may be a subset of T cells or a third type of lymphocyte.

The importance of this type of cell injury in man is not known, but there is evidence suggesting that it contributes to the destruction of tumour cells in experimental animals.

Immune complex, Arthus-type (type 3) reactions

These result from formation of immune complexes by union of antigen with free IgG or IgM antibody with consequent activation ('fixation') of complement. This leads, in turn, to acute inflammation with accumulation of polymorphs and aggregation of platelets. The polymorphs phagocytose the immune complexes and release lysosomal enzymes which cause tissue injury and aggravate the inflammatory response directly and by activating the kinin, clotting and plasmin systems (Fig. 3.13, p. 55). Depending on the distribution of antigen, the reaction may be localised to a particular tissue, and is then termed an Arthus reaction, or immune complexes may form in the blood, producing a generalised reaction commonly known as 'serum sickness' or circulating immune-complex disease.

The local or Arthus reaction

This was described in 1903 by Arthus, who injected rabbits repeatedly with horse serum. When the animals had developed a high level of circulating antibodies to horse serum proteins, he noticed that a subcutaneous injection of horse serum induced a local acute inflammatory reaction, developing over a few hours and sometimes progressing to necrosis. It has since been shown that the reaction may be induced by local injection of a soluble antigen into various tissues in animals with a high level of the corresponding precipitating antibody in their blood. It can be induced also in animals immunised passively by intravenous injection of precipitating antibody (passive Arthus reaction). Localisation of the reaction depends on precipitation of all the antigen in the tissues around the injection site, and thus on a high titre of precipitating antibody in the plasma.

Histological features. Microscopy of the Arthus reaction shows the typical changes of acute inflammation with congestion of small vessels, inflammatory exudation, and marked paving and emigration of neutrophils.

polymorphs. There may be aggregation of platelets in the small vessels and, depending on the severity, haemorrhages and thrombosis, and necrosis extending from the walls of small vessels to surrounding tissues.

Mechanism. Immunofluorescence techniques have demonstrated precipitates of antigen-antibody complexes in the lesion, particularly in the walls of venules. Fixed components of complement may also be detected in the precipitates. The Arthus reaction is largely suppressed in animals by depletion either of neutrophil polymorphs, e.g. by nitrogen mustard, or of complement, e.g. by cobra-venom factor; the intensity of the reaction is reduced by administration of corticosteroids. From such evidence it has been deduced that the reaction is brought about as follows (see also Fig. 6.6). Immune complex formation and deposition in the walls of venules leads to activation of complement, the products of which include **anaphylatoxins*** (C3a and C5a) which bring about acute inflammation by releasing histamine, etc., from mast cells. C5a, and possibly other complement products, are chemotactic for polymorphs, which consequently migrate into the vessel walls and surrounding tissues in large numbers and phagocytose the immune complexes. In so doing, they release lysosomal enzymes and cationic proteins which cause further tissue damage, digest proteins with production of kinins and other vaso-active peptides, and thus aggravate the inflammatory reaction. Platelet aggregation occurs in the damaged vessels and initiates thrombosis with consequent ischaemic necrosis.

The Arthus reaction in man was commonly seen in the days when crude preparations of horse antitoxic globulin or whole antitoxic serum was administered in the prevention and treatment of diphtheria, tetanus, etc. This resulted in the development of precipitating antibodies to horse proteins and a subsequent subcutaneous or intramuscular injection of horse globulin or serum elicited an Arthus reaction. More recently, it has been shown that the Arthus reaction is the basis of **extrinsic allergic alveolitis**, a good example of which is 'farmer's lung'. The farm worker inhales large numbers of the spores of bacteria growing in mouldy

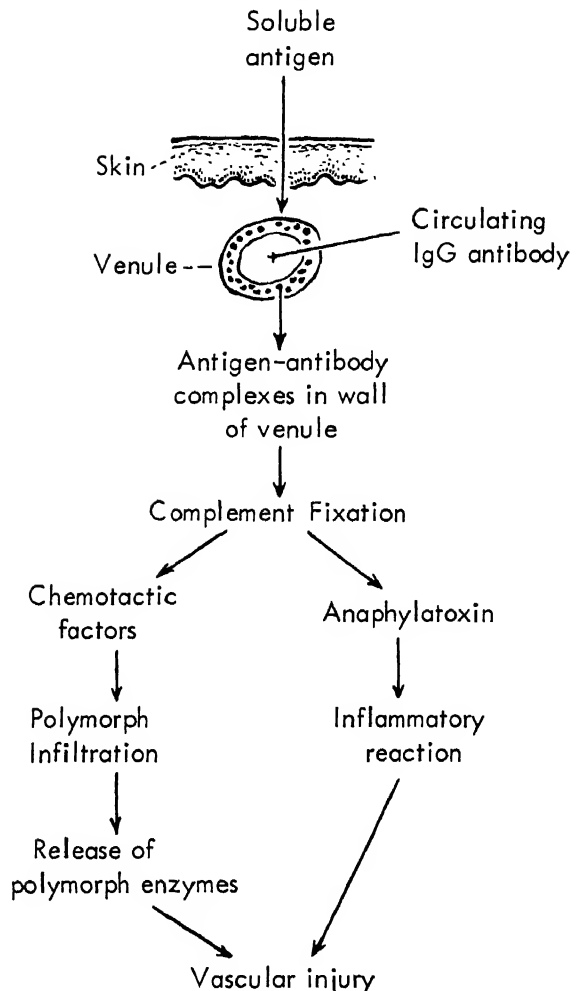


Fig. 6.6 Mechanism of the Arthus reaction. Local injection of soluble antigen into an animal with a high plasma level of the corresponding antibody of IgG class results in union of antigen and antibody in the walls of venules. Complement reacts with the antigen-antibody complexes, and reaction products of complement induce inflammation and polymorph infiltration. The release of polymorph enzymes brings about vascular and tissue injury.

hay; bacterial antigen is absorbed via the alveolar walls and stimulates antibody production. Subsequent inhalation of the spores induces an acute Arthus reaction in the alveolar walls. It has become apparent that individuals with precipitating antibody in their serum vary greatly in their susceptibility to farmer's lung: there is some evidence suggesting that IgE antibody is also necessary, and that the Arthus reaction is

* The term *anaphylatoxin* is unfortunate, because products of complement fixation do not participate in classical anaphylaxis in man, which is due to IgE antibodies.

triggered off by an initial atopic reaction which allows escape of antibodies of IgG class into the vessel walls. As in the rabbit, Arthus reactions in man are inhibited by administration of glucocorticoids.

Circulating immune-complex disease: serum sickness

In man, antigen-antibody complexes are formed in the plasma, both as a result of administration of foreign proteins and haptenic drugs, and also in a number of natural diseases, particularly infections. Serious effects result from their deposition in the walls of blood vessels, especially in the glomeruli, but also in the skin and the walls of arteries. Local lesions develop at these sites and, depending on the duration of deposition, may be acute and self-limiting, recurrent or chronic.

Experimental basis. The basis of this form of hypersensitivity has been elucidated by Dixon, Cochrane and others (see Cochrane, 1973), mainly in rabbits. After a single injection of a large amount of antigen, e.g. bovine serum albumin, no harmful effects occur until, after several days, antibody is produced. As it appears, it combines with antigen still present in the plasma, forming immune complexes. Initially antigen is present in relative excess and its union with antibody produces small soluble complexes (Fig. 5.6, p. 110) which are not readily phagocytosed and so persist in the circulation. As antibody increases, intermediate-sized soluble, and then large, insoluble complexes are formed, and after a few days free antibody can be detected. The larger aggregates of immune complex are rapidly taken up and destroyed by phagocytic leukocytes and by macrophages in the liver, spleen, etc. Accordingly, in a rabbit producing a lot of precipitating antibody, complexes disappear from the plasma in a few days (Fig. 6.7). During this period, however, in which soluble complexes formed in antigen excess are present in the circulation, their presence triggers off a series of reactions leading to release of histamine and other vasoactive agents, with consequent increase in vascular permeability. This in turn allows the soluble complexes, along with plasma proteins, to leak out between the endothelial cells of various blood vessels. At the sites of leakage, complexes are trapped and

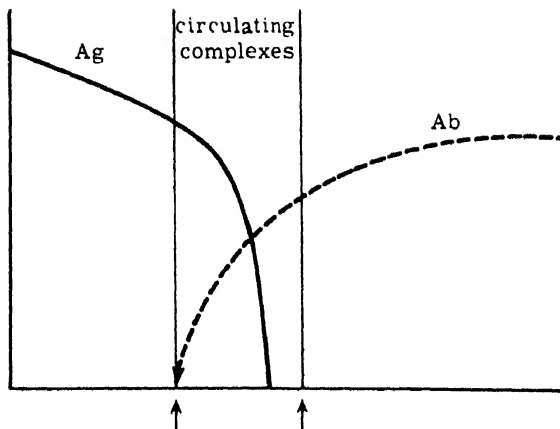


Fig. 6.7 The formation of antigen-antibody complexes in the circulation. Injection of antigen is followed after some days by the appearance of antibody in the plasma: during the next few days, the concentration of antigen falls sharply and antigen antibody complexes are present in the plasma.

accumulate between the endothelium and basement membrane, where their presence results in vascular injury.

The mechanism of increase in vascular permeability, as in inflammation, is complicated, and varies in different species. In the rabbit, in which most of the reservoir of histamine in the blood is in the platelets, histamine release appears to be due mainly to union of antigen with IgE antibody bound to basophil leukocytes (i.e. a type I hypersensitivity reaction): this induces release of a factor which causes the platelets to aggregate and discharge their histamine. Other mechanisms of release of histamine from platelets involve activation of complement by the complexes and participation of neutrophil polymorphs.

In the experiment described above, in which rabbits are given a single injection of an antigen, immune complexes in the blood are deposited beneath the vascular endothelium, particularly in the glomerular capillaries, the small vessels in the joints and skin, in the endocardium, and focally in various arteries. The deposited complexes continue to fix complement, and, except in the glomeruli, this triggers off an Arthus-type reaction, with acute inflammation, infiltration of polymorphs, and sometimes thrombosis and necrosis. The reaction is particularly intense in the arterial lesions, which may extend to involve the whole thickness of the wall. Activation of complement and polymorphs results in phagocytosis of the deposited complexes within 48 hours, so that, although

intense, the reaction is brief. Like the local Arthus reaction, it can be suppressed by prior depletion of polymorphs or complement. In the glomeruli, complexes deposited within or on the epithelial side of the capillary basement membrane fix complement, and yet inflammation is relatively mild and polymorphs are scanty: an obvious explanation is the one-way flow of filtrate from the capillary lumen to the urinary space, which presumably washes away the anaphylatoxins and chemotaxins produced by complement activation. Complexes persist in the glomeruli for many days, and cause glomerular injury by some unexplained mechanism, with resultant proteinuria.

Chronic immune-complex disease can be produced in rabbits by giving daily intravenous injections of soluble antigen in amounts sufficient to provide a period of relative antigen excess over antibody in the plasma after each injection. In contrast to acute serum sickness, the complexes are deposited solely in the glomerular capillaries, where they give rise to lesions resembling various forms of progressive glomerulonephritis in man. Deposition of immune complexes is also influenced by the nature of the antigen and quality of the antibody. Unless there is gross antigen excess, precipitating antibody forms large insoluble complexes which, as stated above, are rapidly phagocytosed and cause little or no injury, so that antigen excess is necessary for formation of smaller, pathogenic complexes. Antibodies which can only react with very few determinant sites on the antigen molecule and have poor avidity (p. 107) also form soluble complexes, even when present in relative excess. Accordingly, some antigen-antibody complexes can produce lesions even when there is a relative excess of antibody in the blood.

Circulating immune-complex disease in man results from injection of heterologous immunoglobulin (now an uncommon cause) and administration of potentially haptenic drugs. It also occurs naturally in systemic lupus erythematosus, in which auto-immune complexes are formed, and in various infections. The formation of immune complexes in the blood may produce both a *general reaction*, and lesions in the glomeruli and elsewhere resulting from *complex deposition in the walls of blood vessels*.

The acute generalised reaction. In its extreme form, for example when a large amount of

foreign protein is injected into an individual with a high titre of the corresponding antibody in the plasma, the rapid formation of high concentrations of immune complexes in the plasma may induce a rapid collapse: this is attributable to intense activation of complement, release of polymorph lysosomal enzymes, etc., and activation of the kinin, clotting and plasmin systems. The clinical features are similar to those of anaphylactic shock (p. 146). When a foreign protein is injected for the first time, classical serum sickness develops 7–10 days later, when production of antibody results in the formation of immune complexes in the plasma. It is a short febrile illness characterised by intense itching of the skin and urticaria, swelling of peripheral joints, and enlargement of lymph nodes. Histamine antagonists bring partial relief. Examination of the serum reveals abnormally low levels of complement components and the presence of products of complement activation. Immune complexes are also demonstrable, although the available techniques are not entirely satisfactory. The polymorph count is at first low, but later raised. After recovery, free antibody appears in the serum. These features indicate that immune complexes form in the blood, activate complement, and trigger off the release of vaso-active agents, including histamine. Phagocytosis of circulating complexes probably results in degeneration and disappearance of most of the polymorphs, and also in the release of endogeneous pyrogen and thus the development of fever (p. 189).

The mechanism of release of histamine, etc. involves the anaphylatoxins of complement, release of polymorph lysosomes, activation of Hageman factor and production of kinins. It may also be that, as in the rabbit, basophil polymorphs or mast cells sensitised with IgE antibody are involved. This possibility is supported by the frequency of bronchospasm suggestive of atopy, and by the occasional severe circulatory collapse as in generalised anaphylaxis.

The typical attack of serum sickness which followed injection of crude antisera is now uncommon, but similar reactions can follow administration of drugs which can confer antigenicity to plasma proteins, and the severe form of dengue fever is due largely to circulating virus antigen-antibody complexes. In spirochaetal infections, including syphilis, and in lepromatous leprosy and some other chronic bacterial infections, the first dose of treatment

by an effective drug may kill very large numbers of micro-organisms and so release microbial antigen, which, depending on the level of circulating antibody and the amount of antigen released, either forms complexes in the blood or induces local Arthus reactions in the lesions. The features of the reaction (*the Jarisch-Herxheimer reaction*) suggest that both phenomena may occur.

Immune-complex deposition is an important cause of glomerulonephritis in man. The typical acute glomerulonephritis following a streptococcal throat infection resembles that of acute serum sickness in the rabbit. It develops when antibody to streptococcal antigen enters the blood and immune complexes are formed and deposited in the glomerular capillaries. Other infections and drug hypersensitivities can have the same effect, and more chronic glomerular injury occurs from prolonged or intermittent immune-complex formation in quartan malaria, lepromatous leprosy and some other chronic infections. In systemic lupus erythematosus, auto-antibodies develop which react with various cellular constituents, e.g. DNA, and complexes formed in the blood are deposited in small vessels in the skin, glomeruli and elsewhere. In most types of human immune-complex glomerulonephritis, however, the nature of the antigen is unknown: the various patterns of disease depend partly on the size of the circulating complexes and the duration and rate of their deposition, and there is also evidence that antigen may be deposited in the glomerular capillary walls, followed by the binding of circulating antibody to form complexes.

Arterial lesions due to complex deposition are less common in man, and their nature is usually difficult to prove because, as in the rabbit, the immune complexes are phagocytosed rapidly by polymorphs. Nevertheless, the focal lesions of polyarteritis nodosa and some other forms of arteritis appear to be of this nature, and surface antigen of the hepatitis B virus has been implicated in some cases.

In many patients with immune complex disease, IgM antibodies develop which are capable of reacting with IgG. These 'rheumatoid factors' react most avidly with the IgG in immune complexes, and may cause circulatory disturbances by increasing the viscosity of the blood, or aggravate the injury caused by deposited immune

complexes. The IgM-IgG complexes often precipitate in cooled serum, and are then termed cryoglobulins.

The lesions produced by immune complex deposition in the kidneys and elsewhere are described in more detail in the appropriate chapters.

Delayed hypersensitivity (DHS or type 4) reactions

Antibody production and cell-mediated immunity are both parts of the normal immune response to most antigens. The union of antibodies with antigens can result in the hypersensitivity reactions described above. Delayed hypersensitivity (DHS) does not involve antibody, but is mediated by the specifically-primed T lymphocytes produced in the cell-mediated immune response. By means of their specific surface receptors, these cells can bind to the antigen which has stimulated their production, and this results in tissue injury characterised by a slowly developing inflammatory reaction—hence *delayed* hypersensitivity.

The reaction of primed T lymphocytes with microbial antigens is an essential defence mechanism against many pathogenic bacteria, viruses and fungi, etc.: the DHS reaction promotes their destruction, and the accompanying tissue injury is the price which must be paid for this protection. Cell-mediated immunity to tumour-cell antigens is known to develop in some cancer patients, and is being intensively investigated in the hope that it may be utilised for the effective destruction of tumours by DHS reactions.

Cell-mediated immunity develops also against harmless foreign antigens, body constituents modified by foreign haptens, against transplanted allogeneic cells or tissues, and sometimes even against the individual's own apparently normal tissue cells. In these circumstances, unwanted DHS reactions occur, resulting respectively in contact dermatitis, rejection of transplants, and auto-immune disease.

Morphological features

The DHS reaction can occur in any part of the body where primed T lymphocytes encounter the corresponding antigen. Its induction in the

skin is used as a test for cell-mediated immunity to various antigens, the classical example being the **tuberculin reaction** in which a small amount of tuberculo-protein ('purified protein derivative' or PPD) is applied to the skin or injected intradermally as in the Mantoux test. This has no effect in non-immune individuals, but in subjects who have developed cell-mediated immunity to tuberculo-protein as a result of tuberculosis or immunisation with BCG (attenuated *Mycobacterium bovis*) the typical delayed inflammatory reaction appears in 12–24 hours and persists for 48 hours or more. The skin becomes reddened and a firm central nodule appears. In a highly sensitised individual, necrosis and ulceration may follow.

Microscopically, the major features are microvascular congestion, accumulation of lymphocytes in and around the small vessels, and swelling of the collagen, apparently due to inflammatory oedema. At the height of the reaction there is intense infiltration with lymphocytes and occasional macrophages, both in and around the capillaries and venules, particularly round the sweat glands and hair follicles (Fig. 6.8). These are the features of the DHS reaction to a soluble protein in man. In animals, accumulation of polymorphs and macrophages, in addition to lymphocytes, is much more prominent. The morphological features of DHS reactions in infections are complicated by the injuries inflicted directly by the micro-organisms or their toxins and by other types of hypersensitivity reactions. A glance at the microscopic appearances of various inflammatory lesions in which DHS is a prominent feature will show considerable differences (e.g. tuberculosis, p. 209; tuberculoid leprosy, p. 215; typhoid fever, p. 10, and contact dermatitis, p. 161). In general, infiltration with macrophages, lymphocytes and lymphoblasts is prominent. The macrophages may also change to epithelioid cells, fuse to form giant cells and undergo necrosis.

The mechanism of DHS reactions

DHS reactions occur when specifically-primed T lymphocytes (memory T lymphocytes) encounter antigen with which they can react. To encounter antigen in the tissues, specifically responsive T lymphocytes must leave the blood in the vicinity of the antigen, and the factors

involved in such emigration are largely unknown.

Circulating T lymphocytes are known to wander through many tissues (p. 121), and there is also evidence that specifically-primed lymphocytes respond chemotactically to the corresponding antigen (Wilkinson *et al.*, 1977), but it is not known whether they are T and/or B cells, nor whether they respond chemotactically *in vivo*. In animal experiments involving passive transfer of radio-labelled lymphocytes, it has been shown that the lymphocytes which aggregate at the site of a delayed hypersensitivity reaction are recently-divided cells and pre-mitotic immunoblasts. In these experiments it was also shown that only a small proportion of the cells accumulating at the test site were specifically primed to the antigen. It still remains undecided whether specifically-primed T lymphocytes are attracted preferentially to the antigen site, and by release of lymphokines (see below) attract other lymphocytes, or whether any recently-divided or stimulated lymphocytes are attracted to the site of injection of any antigen or other, non-specific tissue injury.

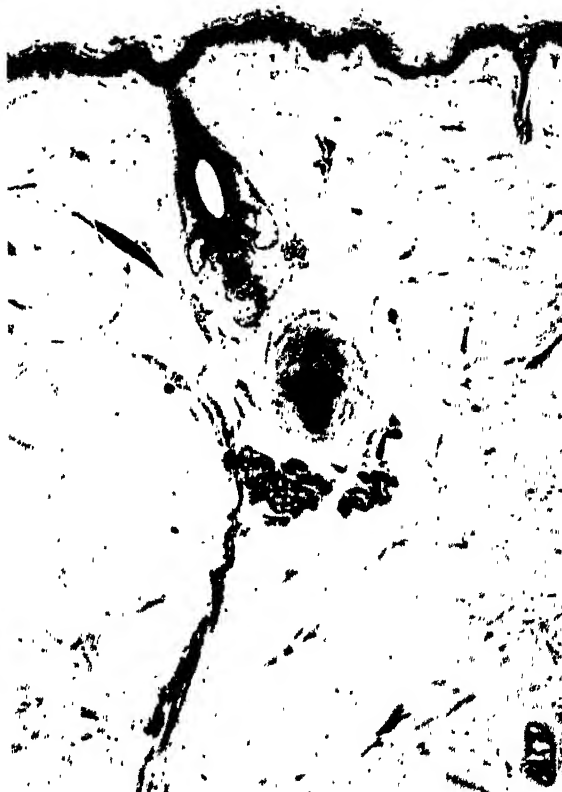
Having encountered and bound antigen by means of its surface receptor, the specifically primed T lymphocyte is stimulated to enlarge and synthesise DNA, transforming into an immunoblast: it also secretes a number of soluble compounds known as *lymphokines*, which are described below. When the antigen is a surface component of a living cell, the T lymphocyte may bind to and kill the target cell: such direct cytotoxicity is of importance in the rejection of allografts (p. 166) and killing of virus-infected cells (p. 195) and tumour cells (p. 315), and probably plays a major pathogenic role in the organ-specific auto-immune diseases (p. 162).

There has been much discussion on the relationship between various functional types of T cells—memory cells, helper and suppressor cells, cytotoxic cells and lymphokine secretors. Until their inter-relationship becomes clearer, it is reasonable to adopt the view, as a working hypothesis, that T memory cells can develop any of these functions.

Lymphokines. As noted above, memory T lymphocytes react with the corresponding antigen and secrete soluble compounds termed lymphokines. The reaction may conveniently be elicited by adding antigen to a suspension of lymphocytes and the supernate examined for lymphokines by a combination of *in-vitro* and *in-vivo* tests. By such means, the following pro-



a



b



c

Fig. 6.8 Positive **a** and negative **b** Mantoux tests. Note the heavy cellular infiltrate around the sweat glands and pilo-sebaceous units in **a**. This distribution is determined by the vascularity of the skin appendages. $\times 43$. At higher magnification **c**, the infiltrating cells are seen to be mostly lymphocytes, which are aggregated around the small blood vessels. $\times 470$. (The late Dr. Janet Niven.)

properties of lymphokines have been demonstrated.

1. *Induction of acute inflammation.* Intradermal injection demonstrates a factor which induces congestion of small blood vessels and inflammatory oedema. This could account for these features in the DHS reaction.

2. *Effects on mononuclear phagocytes.* These are complex, but there is evidence for the following.

(a) *A chemotactic factor.* This induces chemotaxis of monocytes or macrophages *in vitro* and emigration of monocytes *in vivo*; it may account for accumulation of macrophages in DHS reactions.

(b) *A macrophage immobilising factor,* demonstrable by its inhibitory effect on the migration of macrophages, e.g. from the open end of a horizontal capillary tube. When the tube is immersed in tissue culture fluid, addition of this factor to the fluid inhibits migration. It may also play a role in the accumulation of macrophages in DHS reactions.

(c) *A macrophage activating factor,* which increases the metabolic activity of macrophages and enhances their mobility and capacity to phagocytose and kill micro-organisms. This microbicidal effect has been demonstrated *in vitro* and *in vivo*.

(d) *A specific macrophage-arming factor (SMAF)* has been demonstrated in experimentally-induced cell-mediated immunity to tumour cells. The DHS reaction between primed T cells and tumour cells releases a factor which confers on macrophages enhanced killing properties specific for the tumour cells. This factor differs from (c) above in that it is antigen-specific.

3. *Other factors.* Another factor released when primed T cells react with antigen is cytotoxic for tissue cells, and may contribute to the necrosis commonly seen in DHS reactions. This is distinct from the antigen-specific killing of target cells by T lymphocytes mentioned above, which requires close contact.

Other lymphokines are chemotactic for lymphocytes and induce mitosis of lymphocytes: these properties may explain why so many of the lymphocytes in DHS reactions are not specifically primed to react with the antigen which has induced the reaction (see above).

Elucidation of these important factors is still at an early stage. Their importance as mediators of the changes seen in DHS reactions is suggested by their detection not only in reactions in test tubes, but also in extracts of DHS reaction sites.

Man and other primates show, in general, much stronger cell-mediated immune responses and DHS reactions than do lower animals. It is therefore unwise to assume that the findings for guinea-pigs, etc., are applicable to man. Lawrence's transfer factor (p. 131) may be important in cellular immunity and DHS reactions in man, in whom it is far more readily demonstrable, and has a much more lasting effect, than any counterpart in guinea-pigs, etc.

Until recently, it was widely assumed that the presence of lymphocytes in a hypersensitivity reaction was a good indication of a DHS component, but B lymphocytes can and do migrate into sites of antigen (hence the presence of plasma cells in many infections): also it is now known that so-called K cells (p. 152), which have the appearances of small lymphocytes, may bind by surface Fc receptors to target cells sensitised with IgG class antibody and bring about their destruction. The *in vivo* significance of this co-operative cytotoxic effect of antibody and K lymphocytes is not yet known. The presence of macrophages in hypersensitivity reactions is not necessarily indicative of DHS, for they are attracted also by antigen-antibody complexes and by non-antigenic foreign and endogenous material (pp. 61–2, 67).

Systemic effects of DHS reactions

Although this account has concentrated on local DHS reactions, systemic reactions also occur. For example, injection of relatively large amounts of tuberculo-protein into an individual who has developed cell-mediated immunity to it results not only in a severe localised DHS reaction at the injection site, but also fever, malaise and a fall in the level of circulating lymphocytes. If the individual has active tuberculosis, the DHS reaction of the lesion is also aggravated, with extension of necrosis. These effects are known collectively as the *Koch phenomenon* after Robert Koch* who first described them. They are probably attributable to

* The German bacteriologist who, in 1882, discovered the tubercle bacillus and showed it to be the cause of tuberculosis.

the release of lymphokines following the reaction between tuberculo-protein and specifically primed T cells in the blood, lymphoid tissues and tuberculous lesions. The fever of the Koch phenomenon and of active tuberculosis is probably a secondary effect, due to release of endogenous pyrogen (p. 189) by macrophages activated by lymphokines. The Koch phenomenon is not observed in individuals who have not developed cell-mediated immunity to tuberculo-protein.

Hypersensitivity to drugs and chemicals

Most drugs and chemicals which cause hypersensitivity reactions do so because they or their metabolic products combine with host proteins and act as haptens. At first sight, this seems to contradict the observation that haptens can only stimulate an immune response when combined with *foreign* carrier proteins to which the recipient develops cell-mediated immunity (p. 129). The explanation is that the many drugs and chemicals which cause hypersensitivity reactions not only act as haptens, but alter the configuration of the protein molecules with which they combine, thus rendering them 'foreign'.

The type of hypersensitivity reaction which results will then depend on the nature of the immune response, the particular cell or tissue constituent with which the hapten has complexed, the route of administration and dose, etc. In individuals with a tendency to atopy, reaginic antibodies may develop, and further administration of the hapten can then induce a *type 1 reaction*, e.g. asthma or hay fever if the hapten is inhaled as a vapour or airborne suspension, an immediate inflammatory reaction if it is applied locally, or a generalised anaphylactic reaction if a large amount of hapten is absorbed by any route. Anaphylactic reactions to penicillin and related compounds are not uncommon, and have resulted in a number of deaths: in most instances the hypersensitivity is directed towards the penicilloyl degradation product of penicillin. The development of IgG class antibody to a haptenic drug or chemical can give rise to *local reactions of Arthus type (type 3)* when the hapten is localised to one particular area within the tissues, or can lead to *formation of complexes of hapten and anti-*

body within the plasma, with the risk of circulating immune-complex disease (p. 155).

A number of drugs which act as haptens stimulate the production of antibodies which, although not cytotoxic, can bring about destruction of red cells, leukocytes or platelets. For reasons unknown, complexes of these drugs with antibody bind to cells of the blood, and although such binding is often loose, activation of complement by the drug-antibody complex results in C5b and subsequent complement components building up on the cell surface to produce the haemolytic C5b-9 complex (p. 144), with consequent 'bystander' or 'reactive' cytolysis: the binding of immune complexes and complement components also promotes phagocytosis of the cells in the spleen. Drugs which form complexes with these effects include sulphonamides, phenacetin, chlorpromazine and rifampicin, but the list is long and differs for red cells, leukocytes and platelets.

Thirdly, binding of a haptenic drug or drug metabolite to red cells, leukocytes or platelets may render the cells susceptible to injury by antibody to the drug (*cytotoxic (type 2) antibody reaction*—p. 150). It is not known whether tissue cells are injured in this way.

Lastly, cell-mediated immunity may develop towards the hapten-protein complex, and, as described above, subsequent absorption of the haptenic compound gives rise to a *delayed hypersensitivity (type 4) reaction*. This is seen in *contact dermatitis* in which relatively simple chemicals behave as haptens: they are absorbed into the body, often through the skin, and combine with tissue proteins. Cell-mediated immunity develops against the modified proteins, and subsequent skin contact with the same chemical induces a delayed hypersensitivity reaction (Fig. 6.9), causing inflammatory lesions with cellular infiltration, predominantly lymphocytic, and oedema which affects both the dermis and epidermis and progresses to formation of vesicles. The substances capable of inducing this condition are very numerous, and include particularly chemicals which combine firmly with proteins, e.g. dyes, chrome salts, formalin and various derivatives of benzene. Reactions which appear to be of this type occur commonly in women in relation to nickel fasteners on underclothes, some of the nickel being dissolved by acid sweat and absorbed

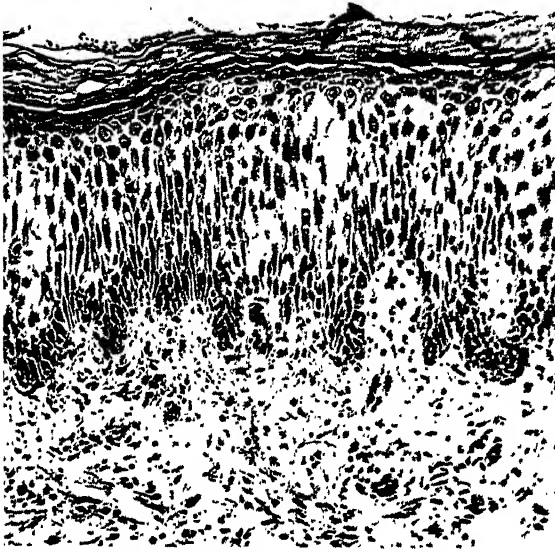


Fig. 6.9 Contact dermatitis. Note oedema of epidermis and perivascular infiltration of lymphocytes in the dermis. The patient had developed hypersensitivity to chromium salts used as a hardener in cement. $\times 110$.

presumably as nickel salts which can combine with skin proteins. Contact with some plants, e.g. poison ivy and primulas is another common cause, and cases also occur from the use of hair dyes, such as paraphenylenediamine, which, of course, bind firmly to keratin. Contact dermatitis results also from application of various medicaments to the skin. It is noteworthy that any individual can be sensitised to various chemicals, and contact dermatitis can thus be induced, but nevertheless some people appear to develop it more readily than others. This is seen in the use of various adhesive surgical dressings which usually produce no reaction but in some instances lead to contact dermatitis, and a similar effect may result from wearing rubber face masks.

It is important to appreciate that the above account is an oversimplification. In many instances, hypersensitivity to drugs or chemicals is extremely complex and the clinical features often conflict with the results of the various available tests for hypersensitivity. With the ever-increasing number of chemicals used therapeutically and in industry, it is not surprising that hypersensitivity reactions, particularly those manifested in the skin and mucous membranes, are becoming increasingly common. Apart from the possession of highly reactive

groups by which they can bind to proteins, it is not yet known what properties of a substance are related to the likelihood of its stimulating the development of hypersensitivity. Individuals predisposed to atopy and patients with systemic lupus erythematosus are prone to develop drug hypersensitivities, but, apart from this, predisposition to drug reaction is quite unpredictable.

Auto-immunity and auto-immune diseases

Some confusion exists over the definitions of auto-antibodies and auto-immune diseases. *Auto-antibodies may be defined as antibodies which react with the individual's own normal body constituents* (which may accordingly be termed **auto-antigens**). This definition excludes antibodies which react only with body constituents which have been altered, for example by a haptenic drug, and so have become 'foreign' to the individual (see above). The definition does not, however, assume that normal body constituents have necessarily stimulated the production of the auto-antibodies. For example *Streptococcus pyogenes* possesses antigens similar to constituents of human myocardium and a streptococcal pharyngitis can induce antibodies which react with normal myocardium (p. 415): these qualify as auto-antibodies. Within this definition, *auto-antibodies to several cell products or constituents are quite commonly present in the serum of individuals both with and without clinical evidence of disease*: they include, for example, antibodies to thyroglobulin, to thyroid epithelial cells, to gastric parietal cells and to the deoxyribonucleoprotein of cell nuclei. These antibodies all react *in vitro* with the individual's own body constituents and with those obtained from others, so that they are acceptable as true auto-antibodies.

Although auto-immunisation occurs without clinical disease, it is strongly associated with a number of diseases. For example, most apparently normal individuals with thyroid auto-antibodies have been shown to have sub-clinical chronic thyroiditis, and patients with more severe, clinically apparent chronic thyroiditis usually have high titres of thyroid antibodies. Similarly, high titres of antibodies to deoxyribonucleoprotein occur especially in patients with the connective tissue diseases, and particularly in systemic lupus erythematosus.

Cell-mediated auto-immunity has also been demonstrated by in vitro tests in some diseases. For example thyroid auto-antigens inhibit the migration of leukocytes (p. 113) of patients with chronic thyroiditis.

It is thus apparent that, in some diseases, there is evidence of auto-immunisation against particular body constituents. These are the so-called **auto-immune diseases**. In most, there is still a lingering doubt that the lesions are due to immunological reactions against the target auto-antigens; in some instances this probability is supported by production of similar lesions by auto-immunisation of animals, or by investigations on naturally-occurring auto-immune diseases in inbred strains of animals.

There are also a number of examples in which auto-antibodies develop as a result of tissue injury and appear to be without pathogenic effect, an example being auto-antibodies to myocardial cells which frequently develop following ischaemic necrosis of the myocardium; presumably antigen is released by the dead muscle cells and stimulates an immune response.

Auto-immune diseases may be classified into: (1) a group of organ-specific diseases affecting glandular tissues; (2) systemic lupus erythematosus and possibly the other connective tissue diseases; and (3) a number of miscellaneous diseases which do not fit readily into either of the above classes.

The organ-specific auto-immune diseases

These are characterised by chronic inflammatory destruction of a particular glandular tissue accompanied by the presence in the plasma of auto-antibodies which react specifically with normal cellular components of the target tissue. The tissues affected include the thyroid, gastric mucosa, adrenal cortex, parathyroid glands and cells of the pancreatic islets of Langerhans. The main features are exemplified by **chronic auto-immune thyroiditis**, in which infiltration of the thyroid gland by lymphocytes, plasma cells and macrophages is accompanied by glandular epithelial destruction and fibrosis. These changes may be focal and are then sub-clinical and usually non-progressive, or they may be diffuse, giving rise to either thyroid enlargement, sometimes with hypofunction (*Hashimoto's thyroiditis*), or de-

struction and shrinkage of the gland with gross hypofunction (*primary myxoedema*). Auto-antibodies to normal thyroid constituents are detectable in the serum in virtually all cases of Hashimoto's thyroiditis, and in most cases of primary myxoedema and sub-clinical thyroiditis. They include antibodies reactive with: (1) thyroglobulin, often in sufficient concentration to give a precipitin reaction (Fig. 5.7, p. 110); (2) a second constituent of thyroid colloid; and (3) cell membrane constituents ('microsomes') of thyroid epithelium—the so-called thyroid microsomal antibody (Fig. 5.11, p. 112). Chronic thyroiditis occurs much more often in women than in men, and the incidence increases with age. Over 10 per cent of middle-aged or elderly women have one or more thyroid antibodies and some degree of chronic thyroiditis. There is a general correlation between the presence and titres of the serum antibodies and the extensiveness and activity of the thyroiditis, but the correlation is by no means exact for any one antibody or any combination of antibodies.

As mentioned earlier, a fourth thyroid antibody, which reacts with the TSH receptors of thyroid epithelial cells, is responsible for the hyperthyroidism of Graves' disease which is usually accompanied by focal auto-immune thyroiditis.

Chronic auto-immune gastritis, affecting the acid-secreting mucosa of the gastric fundus, has many points of resemblance to chronic thyroiditis. It affects women more often than men, and the incidence increases with age. In most cases, the serum contains 'microsomal' antibody to gastric parietal cells and, in a minority of cases, antibodies to the intrinsic factor essential to absorption of vitamin B₁₂. The affected mucosa is infiltrated with lymphocytes, plasma cells and macrophages, and all grades of destruction of chief and parietal cells are observed. In most cases, the gastritis is mild and sub-clinical, and progresses very slowly, but in some cases it progresses more rapidly to diffuse atrophy of the mucosa (like the thyroid in primary myxoedema) and parietal cell deficiency then results in achlorhydria and lack of intrinsic factor, the latter leading in some cases to B₁₂ deficiency and pernicious anaemia.

Auto-immune adrenalitis is a much less common condition; it is, however, the major cause of adrenal cortical atrophy and func-

tional deficiency (Addison's disease). The serum commonly contains 'microsomal' auto-antibody to a cell-membrane constituent of adrenocortical epithelium. **Primary hypoparathyroidism** is rare: specific auto-antibodies are demonstrable in the serum in some cases, and the parathyroid glands are shrunken and extremely difficult to find at necropsy.

In addition to their morphological and serological similarities, each of these diseases tends to have a high familial incidence, and moreover the diseases tend to occur in association, both within affected families and in individuals. For example, patients with Hashimoto's thyroiditis have a high incidence of gastric antibody and a particular tendency to develop pernicious anaemia, while thyroid and gastric antibodies, sometimes associated with the corresponding clinical diseases, are unduly common in patients with auto-immune Addison's disease or primary hypoparathyroidism: even these two latter rare diseases have been found to be particularly associated with one another.

Insulin-dependent (type I) diabetes differs from the other organ-specific auto-immune diseases in affecting children and young adults, and there is recent evidence suggesting that it may be initiated by a viral infection. Auto-antibodies to cells of the islets of Langerhans are detectable in the serum in nearly all early cases. The islets at first show lymphocytic infiltration and later become atrophic.

Pathogenesis and aetiology. It has not been proved that these diseases are the result of auto-hypersensitivity reactions, an alternative explanation being that the glandular destruction is due to some other (unknown) agent, and that auto-antibodies develop as a secondary phenomenon. In favour of an auto-immune pathogenesis, auto-antibodies do not, in general, result from destruction of tissue. For example, thyroid injury by large doses of radioiodine or viral thyroiditis does not stimulate the production of thyroid antibodies, although if antibodies are already present their titres may increase: nor does recurrent alcoholic gastritis result in gastric antibodies. Secondly, organ-specific lesions resembling those of the human diseases, but usually reversible, can be induced experimentally in animals by injections of homogenates of the organ (e.g. thyroid or adrenal) incorporated in Freund's adjuvant. The adju-

vant enhances immune responses, particularly those dependent on T lymphocytes (p. 129) and passive transfer of lymphocytes and antibody suggest that cell-mediated immunity is a more important cause of tissue injury than auto-antibodies in these experimental conditions. There are, however, exceptions, one being the chronic thyroiditis which develops spontaneously in an inbred obese strain of chickens: it is accompanied by thyroid auto-antibodies and is more severe in chickens rendered deficient in T cells by thymectomy after hatching, but is prevented by depriving the birds of B cells by early bursectomy (p. 118). The same results apply to artificially-induced auto-immune thyroiditis in ordinary chickens.

In the human diseases, the respective roles of auto-antibodies and cell-mediated immunity in the organ-specific auto-immune diseases are not known. It is noteworthy that in the two examples of diseases in which the function of tissue cells are known to be affected by auto-antibodies (thyrotoxicosis and myasthenia gravis—p. 152), the antigen is a receptor projecting from the surface of the target cell. Thyroid microsomal antibody has been shown, in the presence of complement, to be cytotoxic for thyroid epithelial cells in culture, but only if the cells are first treated with trypsin, and it may be that such treatment is necessary to expose the cell membrane auto-antigen. There is no evidence of thyroid injury in the infants of mothers with Hashimoto's thyroiditis although in some cases high concentrations of thyroid antibodies of IgG class are transferred to the fetal circulation.

It is difficult, in man, to investigate the pathogenic role of cell-mediated auto-immunity, for while macrophage migration inhibition tests (p. 159) suggest that this develops, the evidence is not conclusive, and in any case does not indicate that it causes tissue destruction. The lesions are typically infiltrated with lymphocytes, which suggests a delayed hypersensitivity reaction, but could also represent an antibody-dependent (type 2) cytotoxic reaction (p. 152). Also, there are usually some, and often many, plasma cells in the lesions, and locally produced antibody might have a cytotoxic effect. In spite of the rather flimsy nature of the evidence, there is a fairly general belief that the lesions of these diseases are mediated largely by delayed hypersensitivity reactions.

The familial incidence of the organ-specific autoimmune diseases suggests a genetic predisposing factor, and this is supported by the greater concordance in monozygotic than in dizygotic twins (i.e. if one twin has the disease, the other is more likely to develop it if they are monozygotic). The studies on twins indicate that there must also be environmental predisposing factors, and these have been the subject of speculation, but with little advance. One thyroid auto-antigen, thyroglobulin, is normally present in low concentration in the plasma, and appears to induce 'low-dosage' tolerance of T cells (p. 133); potentially responsive B cells have been demonstrated in normal individuals, but probably they require T-cell co-operation to produce antibody. If this applies to the other auto-antigens in these diseases, then breakdown of T-cell tolerance must be necessary for cell-mediated auto-immunity and auto-antibody production. It has been suggested that T-cell tolerance might be broken by modification of cell constituents by drugs, by microbial products, or by disorders of metabolism, but so far there is no evidence for any of these possibilities in the organ-specific autoimmune diseases. Another possibility is the defective functioning of suppressor T cells (pp. 130, 134).

The connective tissue diseases

Systemic lupus erythematosus. This is one of the so-called connective tissue diseases. It is characterised by acute and chronic inflammatory lesions in many organs and tissues, and by the occurrence in the plasma of various auto-antibodies, most of which react with normal constituents common to most types of cell in the body. The sites of lesions include the skin, muscles, joints, glomeruli, heart and blood vessels, but the distribution varies greatly and may be even wider. Auto-antibody to deoxyribonucleoprotein is nearly always demonstrable in the serum by immuno-fluorescence tests (Fig. 6.10), and antibodies to DNA, RNA and various cytoplasmic cellular constituents are commonly present. There may also be cytotoxic auto-antibodies to red cells, platelets and leukocytes and antibodies to clotting factors in the plasma.

The auto-antibodies to nuclear and cytoplasmic constituents are not cytotoxic, and many of them occur in other diseases and, usually in low titre, in some normal subjects. The corresponding auto-antigens may, however, be released by breakdown of cells, and immune-complex (type 3) reactions can then

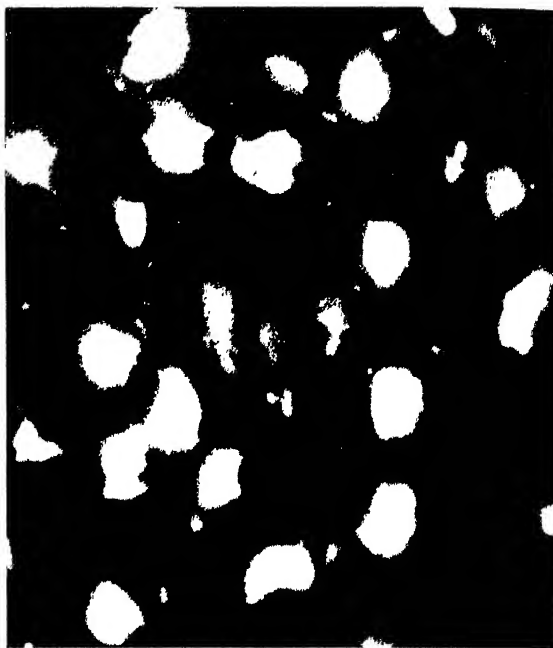


Fig. 6.10 Antibody to deoxyribonucleoprotein demonstrated by the immunofluorescence technique. Note the diffuse nuclear fluorescence. $\times 775$. (Professor J. Swanson Beck.)

result. In fact, most of the pathological features of SLE can be explained on the basis of deposition of circulating immune complexes in the walls of small blood vessels. Disease activity correlates fairly closely with the concentration of anti-DNA in the plasma (measured by DNA binding capacity of the serum) and low levels of serum complement. The glomerular lesions (p. 813) are due to deposition of immune complexes in the glomerular capillary basement membrane and subsequent complement fixation, and this process is responsible also for at least some of the lesions in the skin and elsewhere. Auto-antibodies to native (double-stranded) DNA are virtually specific for SLE and their complexes with DNA contribute significantly to the renal lesions. Antibodies to denatured (single-stranded) DNA occur also in other diseases and do not correlate so closely with disease activity.

The familial occurrence of SLE, and of the various auto-antibodies associated with SLE, raises the possibility of genetic factors, and the spontaneous development of a very similar disease in the F1 hybrid of the NZB and NZW inbred strains of mice developed in New Zealand by Bielchowsky, provides a genetically-determined model.

There is also circumstantial evidence that a virus infection may be involved in SLE. Virus-like particles have been observed in the renal lesions, and are regularly discernible in the tissues of NZ mice. When maintained in tissue culture, lymphoid cells from these mice release C type RNA virus particles which *in vivo* stimulate the development of antibodies reactive with host DNA and RNA. It is possible that this virus, which is responsible for lymphoid neoplasia in mice, modifies the lymphoid cells in some way which predisposes them to auto-immune responses. There is also evidence that the SLE-like disease of NZ hybrid mice is related to thymic deficiency: 'thymic hormone' levels (p. 118) fall at a relatively early age in these mice (and are low in patients with SLE), and the disease in mice is accelerated by neonatal thymectomy and inhibited by injection of thymocytes from young mice. These observations, and the demonstration that animals rendered immunologically tolerant to an antigen develop suppressor T cells which are capable, on transfer to a normal animal, of inhibiting the immune response to that antigen (p. 134), raise the possibility that auto-immune responses are normally inhibited by suppressor T cells, and that auto-immune diseases arise particularly in individuals with T-cell deficiencies. In SLE, which occurs most commonly in women of reproductive age, serum antibody levels to some viruses are unusually high, and cell-mediated immune responses are impaired, suggesting that suppressor T-cell function may be deficient. It is therefore of considerable interest that antibody cytotoxic for a subset of T lymphocytes has been demonstrated in the serum of patients with SLE and occurs also in their close household contacts, including spouses (suggesting a causal viral infection?). The possibility that this auto-antibody destroys suppressor T cells could explain the large numbers of auto-antibody responses observed in this disease. Antibody claimed to be cytotoxic for suppressor T cells has also been detected in NZ mice.

Rheumatoid arthritis (RA). Because of its high incidence and disabling effects, this is the most important of the connective tissue diseases. The main feature is a destructive polyarthritis, in which the synovial membrane is infiltrated with lymphocytes, macrophages and plasma cells. Immune complexes and activated complement components are present in the synovial fluid and are deposited in the synovial membrane. In most cases, the serum contains **rheumatoid factors**: these are immunoglobulins (usually IgM) which behave as antibodies to auto-antigenic components of IgG. Rheumatoid factors react only weakly with native IgG, but strongly with IgG which has been heat-

denatured, and with IgG antibody coupled with the corresponding antigen. Experimental evidence suggests that rheumatoid factors develop when IgG antibody forms immune complexes: binding with antigen alters the IgG molecule and renders it auto-antigenic.

The aetiology and pathogenesis of RA are discussed on p. 939. It appears that inflammatory changes are brought about as a result of activation of complement by antigen-antibody complexes. Initially such complexes might be provided by antibody reacting with a postulated infective agent. Subsequently, complexes are formed by IgG rheumatoid factor which reacts with its own Fc component (Fig. 23.70, p. 939). RA may thus be an Arthus (type 3) reaction, but there are other possibilities, including a delayed hypersensitivity (type 4) reaction between specifically primed T lymphocytes and synovial lining cells.

Polyarthritis is not uncommon in SLE, but is seldom so severe or destructive as rheumatoid arthritis. This and other associations do not indicate an auto-immune pathogenesis for rheumatoid arthritis, but merely suggest that common genetic and possibly environmental factors predispose to both conditions.

The other connective tissue diseases are dealt with in the relevant systematic chapters. Apart from the variable occurrence of anti-nuclear and other auto-antibodies, there is little evidence to suggest an auto-immune pathogenesis.

Other auto-immune diseases

Auto-immune destruction of red cells, leucocytes and platelets by cytotoxic antibodies (p. 150) may occur in isolation, or in association with systemic lupus erythematosus. The pathogenic effects of auto-antibodies in *Graves' disease* and *myasthenia gravis* have already been mentioned (p. 163).

Other diseases in which auto-immunity may be significant include *ulcerative colitis*, in which the intestinal epithelium may be the target cell of an auto-immune response and some types of chronic liver disease, notably *primary biliary cirrhosis* and virus-negative *chronic active hepatitis*, in which there is evidence of auto-immunity to components of bile-duct epithelium and hepatocytes respectively. Various other diseases could be mentioned, but as the list lengthens, the evidence becomes progressively weaker.

Rejection of transplanted tissues

The treatment of burns by skin grafting is a well-established procedure. The epidermis of *autologous grafts* extends to cover the denuded area and survives indefinitely, while *allogeneic grafts* become established, but invariably undergo necrosis within two or three weeks. Evidence that this rejection process is mediated by an immunological reaction on the part of the host was first provided by Gibson and Medawar in 1943. In a series of important experiments, Medawar and his colleagues went on to lay the foundations of transplant immunology. They showed that skin grafts between syngeneic* mice were accepted permanently, while allogeneic grafts stimulated an immune response in the host and were consequently destroyed ('rejected') 1–3 weeks after grafting. They also showed that mice injected at birth with allogeneic cells would subsequently accept permanently a skin graft from the same donor strain and that this state of unresponsiveness—the first experimental demonstration of acquired immunological tolerance (p. 132)—could be abolished, with consequent rejection of the skin graft, by injection of host-strain lymphocytes from a normal mouse or from one that had previously rejected a graft from the allogeneic strain. Lymphocytes from the latter mouse induced more rapid and intense graft rejection, showing that, as a result of previously rejecting an allograft, it had developed persistent immunity, manifested by the reactivity of its lymphoid cells. This early work suggested the importance of cell-mediated immunity in allograft rejection. It is true that antibodies also developed in the recipients of allografts, but their injection into tolerant animals bearing an appropriate allograft did not result in rejection.

Medawar's major observations and conclusions have been confirmed in experiments involving transplantation of various tissues in many vertebrate species. The mechanism of rejection is complex, but in most situations cell-mediated immunity plays a major role and the graft is destroyed mainly by a delayed hypersensitivity (type 4) reaction. Specifically-primed T lymphocytes bind to 'transplant' antigens on the surface of the graft cells (see below)

and bring about their destruction. The mechanism of such cytotoxic activity is not known. The specifically-reactive T lymphocytes also release lymphokines (p. 157), which activate macrophages, and possibly also a specific macrophage-arming factor (SMAF) which enables macrophages to bind specifically to the transplant antigens of the graft cells and destroy them. In organs such as the kidney, which are transplanted by connecting the major blood vessels of the graft to host vessels, injury may result also from a cytotoxic antibody (type 2) reaction (p. 150).

Transplant antigens: the HLA system. The cells of probably all vertebrates possess numerous surface iso-antigens, termed transplant antigens. In several species, including man, a major system of strong antigens, characteristic for each species and determined by multiple alleles at a complex locus, has been demonstrated. In man, this is the HLA system of antigens, so called because the antigens were first detected in human leukocytes. Each individual inherits HLA antigens from each parent as shown in Fig. 6.11.

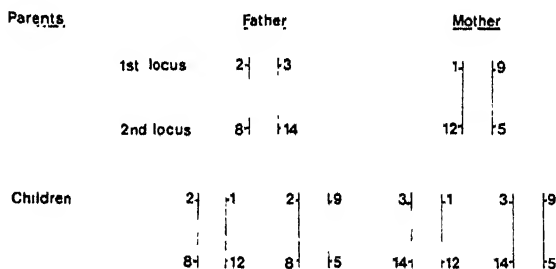


Fig. 6.11 The major histocompatibility complex in man contains a number of allelic genes, at loci close together on the same chromosome (No. 6). The diagram shows the mode of inheritance of genes of the first two loci, which code respectively for one antigen of the HLA-A series (antigens 1, 2, 3, 9, 10, 11, 28 and 29) and one antigen of the HLA-B series (antigens 5, 7, 8, 12, 13, 14, 18 and 27). The individual thus inherits a pair of antigens (one from each parent) in each series. Because of their close linkage, the genes, and so the antigens, are inherited in 'sets', so that siblings with the same antigens in one series (e.g. the same HLA-A antigens) are likely to have the same antigens in the other series. This is shown above for HLA-A and HLA-B antigens, but it applies also to the HLA-C and -D antigens, and explains why there is a 25 per cent chance that two siblings will inherit identical sets of antigens. Transplantation between such siblings is usually highly successful.

* For terminology, see p. 105.

In rats and mice, the major transplant antigens are the main factors in determining the intensity of the iso-immune response, and thus of rejection, following transplantation. In human renal transplantation, matching has so far been limited by the availability of suitable antisera and has been restricted largely to HLA-A and -B antigens which (together with HLA-C antigens) are detectable by means of antisera (see below). Another group of HLA antigens (the D or LD antigens) determined by alleles at a fourth locus were not detected by antisera (although these are now becoming available) but by the mixed lymphocyte reaction, in which a mixture of lymphocytes from two individuals is incubated and a positive reaction is indicated by DNA synthesis and 'blast' cell transformation. Such transformation represents a reaction of T lymphocytes induced by HLA-D antigens on the B lymphocytes (of the other individual). By treating the lymphocytes of one individual with mitomycin before mixing, they are rendered unresponsive and any reaction then represents the response of the T cells of the other individual (the 'one-way' mixed lymphocyte reaction). As explained in Fig. 6.11, the close proximity of all the HLA loci on the chromosome 6 ensures that siblings who have inherited the same HLA-A and -B antigens will also almost always have the same C and D antigens, and will thus be closely histocompatible. By contrast, it is very difficult to find complete HLA identity among unrelated individuals.

Because it codes for strong histocompatibility antigens, the HLA region on chromosome 6 has been termed the *major histocompatibility complex* (MHC). Similar complexes are found in mice and other higher vertebrates. In addition to the HLA (or in the mouse the H2) loci, the MHC contains Ir genes which determine the strength of the thymus-dependent antibody responses to particular antigens: they influence co-operation between T and B cells and also degradation of antigen by macrophages.

Graft versus host reaction. If normal lymphoid cells are injected into an allogeneic host, they will be destroyed unless the host is immunologically deficient or tolerant and cannot mount a rejection reaction. In these circumstances, the grafted cells may survive and mount an immune response against the host, with a

consequent **graft-versus-host (G.v.H.)** reaction. This occurs when allogeneic lymphocytes from an adult mouse are injected into neonates or into mice rendered immunodeficient, e.g. by thymectomy and x-irradiation, and when an F1 hybrid mouse is injected with lymphocytes of either parental strain. The G.v.H. reaction is complex and includes splenomegaly, lymph node enlargement, haemolytic anaemia and predisposition to infections. When induced in the neonate, these changes, together with impairment of growth, have been termed *runt disease*. In man, G.v.H. disease may result from infusion of lymphoid or haemopoietic cells into patients with T-cell deficiency, e.g. following intensive cytotoxic therapy for leukaemia: it presents as anorexia, diarrhoea, dermatitis and liver failure.

Human renal transplantation

Many thousands of kidneys have been transplanted in the past few years to patients with irreversible renal failure. The major obstacle is immunological rejection of the graft and, except for transplants between identical twins, it is essential to protect the graft by administration of immunosuppressive drugs such as glucocorticoids, azathioprine or actinomycin C. Initially, high dosage is necessary to prevent acute rejection, but the dosage can gradually be reduced, in some cases to very low levels, without rejection occurring. This indicates that the host becomes increasingly less responsive to the graft antigens. One possible explanation is that the continued release of antigens by the graft, together with immunosuppression, results in specific immunological tolerance. Another is that the host develops 'enhancing' antibodies (p. 134) which protect the graft by suppressing the cell-mediated immune response or by combining with graft antigens and so concealing them from specifically responsive T lymphocytes. Administration of iso-antibodies reactive with antigens of the grafted tissue has been shown to prolong graft survival in animals, and there is preliminary evidence that administration of HLA antibodies may have a similar enhancing effect in human renal transplantation, although such antibodies (presumably in larger amounts) can also cause immediate rejection (p. 846).

While on large doses of immunosuppressive

drugs, transplant patients are very liable to develop infections, both with virulent pathogens and with opportunistic micro-organisms such as *Cytomegalovirus*, *Pneumocystis carinii* and various fungi.

In spite of these problems, approximately 50 per cent of the transplants have done well and are functioning some years later.

The pathological features of rejection of renal transplants are described on p. 846.

Tissue typing. At present, tissue typing is usually performed by a cytotoxicity test, using typing antisera and complement, upon cells of the individual to be typed: if the cells possess the corresponding iso-antigen, they will be killed (Fig. 2.2, p. 10). All tissue cells, leukocytes and platelets (but not red cells) possess HLA antigens, and it is convenient to use blood lymphocytes for typing. HLA antisera are obtained from recipients of blood transfusions or previous transplants, or from parous women, some of whom have developed antibodies to HLA antigens of the fetus. By testing with a panel of lymphocytes of known HLA phenotypes, and suitable absorption to remove unwanted antigens, specific HLA antisera can be provided.

There is no doubt that renal allografts have a better chance of surviving when the donor and recipient are closely matched for HLA antigens. The effect is greatest in grafting between siblings of identical HLA-A and -B antigens because, as explained on p. 167, they are also very likely to have identical C and D antigens, and over 90 per cent of kidneys are functioning well one year later. Even with close matching between unrelated donors and recipients, the figure is only about 60 per cent because identity at all four HLA loci is unlikely. Testing for histocompatibility of HLA-D antigens by the mixed lymphocyte reaction takes 4 days or so, and is usually impracticable. In addition to the HLA antigen, there are loci on other chromosomes determining weaker transplant antigens, and even grafts between HLA-identical siblings will be rejected unless the recipient receives immunosuppressive therapy.

For reasons explained on p. 846, it is essential to ensure ABO compatibility in human renal transplantation, and to test the recipient's serum for cytotoxic antibody to donor cells.

Transplantation of other tissues

Infusion of **haemopoietic cells**, which include pluripotent stem cells (p. 118), is a logical treatment for infants with congenital deficiency of haemopoietic stem cells, for patients with certain forms of aplastic anaemia, and for children treated for acute leukaemia by cytotoxic drugs in doses which destroy their haemopoietic and lymphoid cells. The depressed immune responsiveness of such patients reduces the risk of rejection of the donated cells, but immunocompetent T lymphocytes in the donation are very liable to respond to transplant antigens of the host, causing a fatal *graft-versus-host reaction* (see above). It is thus essential to use as donor an HLA-identical sibling.

Successful **corneal allografting** has long been practised without immunosuppression of the recipient. This is because the cornea is avascular and therefore a 'protected site' in which the graft does not induce an immune response in the recipient. If, as sometimes happens, blood vessels extend into the grafted cornea, then rejection occurs.

In allografts of **blood vessels** and **tendon**, the cells either die from ischaemia or are destroyed by a rejection reaction, but the collagen and elastic fibres persist, and are repopulated with host cells and vessels: thus the use of stored vessel or tendon is equally, if not more effective. Similarly, the cells of **bone grafts** die, but the matrix may provide the desired mechanical effect (p. 875). The use of **cartilage grafts** in plastic surgery is of considerable interest: both the cells and the matrix of allografts may survive for long periods without inducing a rejection reaction. This is due to the avascular nature of cartilage, and to the matrix which allows diffusion of nutrients and metabolites between graft chondrocytes and host, but acts as an 'immunological barrier' between them.

Another exception to the general phenomenon of allograft rejection is provided by nature's allograft, **pregnancy**. The trophoblast, of fetal origin, is bathed in maternal blood, and yet it is tolerated for nine months, in spite of the presence of incompatible (paternal) transplant antigens in the fetal cells. There is slight depression of maternal immune responsiveness during pregnancy, but the mother does not develop specific immunological tolerance to the fetus. The most likely explanation of failure to

reject the fetus appears to be that the cells of the syncytiotrophoblast are coated with a layer of mucopolysaccharide, which provides an 'immunological barrier'. As already mentioned,

pregnancy commonly results in the development of maternal antibodies to the transplant antigens of the fetus, but this is without apparent effect.

Immunological Deficiency States

There are a large number of conditions in which the normal defence mechanisms against invasive micro-organisms are impaired. For most purposes it is useful to classify such deficiencies into two major groups. Firstly, deficiencies of non-specific resistance, as in diabetes mellitus, impaired function of neutrophil polymorphs (p. 513), etc. This miscellaneous group is dealt with under the appropriate diseases: it includes also lesions which impair resistance locally, for example obstruction of hollow viscera, as in the urinary tract or air passages, and ischaemia of the lower limb leading to gangrene.

In the second major group, which is discussed here, impaired resistance is due to defects in specific immunological responsiveness. These are best classified into primary and secondary types, and also in relation to the type of immunological defect present.

In the group of *primary conditions*, the immunological deficiency becomes manifest usually, but not always, in early childhood, and in most of the conditions there is good evidence that the defect is genetically determined. Other abnormalities, e.g. thrombocytopenia in the Wiskott-Aldrich syndrome, or hypoparathyroidism in the DiGeorge syndrome, may accompany the immunological defect, giving rise to characteristic disease complexes, but the immunological deficiency is not secondary to the other parts of the syndrome. By contrast, the *secondary immunological deficiencies* occur at any age, are not genetically-determined, and the immunological defects are the result of injury to the lymphoid tissues, either by various disease processes, particularly lymphoid neoplasia, or by immunosuppressive agents.

The type of immunological defect present determines the clinical picture and form of therapy required. The division of lymphoid cells into two major classes—thymic-dependent (T) and thymic-independent (B)—has been dealt

with in Chapter 5. Its validity in man is demonstrated by the occurrence of immunodeficiency states affecting mainly T-cell function, with depression of cell-mediated immune responses, or mainly B-cell function, with depression of antibody production: combined deficiencies, affecting both T- and B-cell function, are also observed. Such well-defined immunodeficiencies occur as primary, congenital defects, but they are rare. Less serious and less clearly-defined deficiencies also occur, both as congenital and acquired conditions.

Primary immunological deficiencies

(1) Deficiency of B-cell function. This group is exemplified by **infantile sex-linked agammaglobulinaemia**, which was the first to be described and is sometimes known as the **Bruton type of agammaglobulinaemia** after its discoverer. The major abnormality is a virtually complete inability to produce the three major classes of immunoglobulins—IgG, IgM and IgA. In consequence, there is little or no antibody production in response to infections or immunisation procedures, and the normal blood group iso-antibodies are usually not detectable. The condition is observed nearly always in boys, being transmitted by a gene defect in the X chromosome (sex-linked recessive). Symptoms usually arise in the second year of life, protection before that being provided by maternal antibodies of IgG class transmitted to the fetus. The defect results in unusually frequent and serious bacterial infections, particularly those due to the pyogenic bacteria, and including respiratory and pulmonary infections, meningitis and septicaemia. 'Opportunistic' infections (p. 174), e.g. pneumonia due to the protozoon, *Pneumocystis carinii*, also occur, and candidiasis is common. The infections respond to appropriate antibiotics, and

diagnosis depends upon the demonstration of the near-absence of serum IgG (below 0.5 g/litre), IgM and IgA (below 3×10^{-2} g/litre). Deficiency of IgG cannot be demonstrated until the maternal IgG has fallen to a low level—usually by about 8 months old, although very low levels of the other two immunoglobulins are observed before this time, since they do not cross the placenta.

The lymph nodes and tonsils are small, and biopsy reveals an absence of germinal centres and plasma cells, while rectal biopsy reveals absence of plasma cells in the mucosa. The blood lymphocytes are not greatly diminished, but B lymphocytes are virtually absent. The thymus is normal, and cell-mediated immune responses are not impaired. Accordingly, the responses to BCG and vaccinal immunisation are normal and afford protection, and virus infections in general occur with the same frequency and clinical features as in normal children. Chronic polyarthritis, closely resembling rheumatoid arthritis, is of common occurrence.

The effectiveness of regular injections of human IgG in preventing infections has increased the importance of early diagnosis. It is important to distinguish the Bruton type of agammaglobulinaemia, which requires life-long therapy, from **transient hypogammaglobulinaemia**. This latter condition presents similar clinical features and morphological changes in the lymphoid tissues, but is merely a delay, and not a permanent failure, of the capacity to produce immunoglobulins. It is familial, relatively common, and more frequent and severe in infants born prematurely: it affects both sexes, and the defect disappears within the first three years of life. In most cases, severe immunoglobulin deficiency is limited to IgG, and normal levels of IgM and IgA in the serum may help to distinguish it from the Bruton type.

There are a number of less well-defined conditions which appear to fall within this group. In all of them, there is defective production of one or more classes of immunoglobulins, impairment of antibody production, and relatively normal cell-mediated immune responses. The evidence favouring a genetic predisposition, and a particular mode of genetic transmission, varies in the different types. In some forms, the immunological deficiency does

not become manifest until adult life, and yet the disease tends to occur in families, and often exhibits a familial association with other immunological disturbances, e.g. hypergammaglobulinaemia and systemic lupus erythematosus. Some patients with such late-onset immunoglobulin deficiency also develop autoimmune diseases such as pernicious anaemia and connective tissue diseases, but without demonstrable auto-antibodies.

(2) **Deficiency of T-cell function.** An example of this group is provided by the rare **DiGeorge syndrome**, in which there is almost complete failure of development of the thymus and parathyroids from the third and fourth branchial arches.

In those infants who survive the neonatal period, immunoglobulin production appears normal, although antibody responses, at least to some antigens, are impaired, probably because of lack of helper T cells. The lymph nodes contain plasma cells and germinal centres, but the paracortical (thymus-dependent) areas are deficient in small lymphocytes, and the number of circulating lymphocytes, although variable, is low in some cases. The condition may affect infants of both sexes, and there is no evidence for a genetic predisposition. Affected infants can deal perfectly well with pyogenic bacteria, but suffer from 'opportunistic' infections, e.g. by *Pneumocystis carinii* (p. 474) and fungi and also from severe virus infections. Impairment of cell-mediated immunity is demonstrable by failure to develop contact hypersensitivity to agents such as dinitrochlorobenzene (p. 137) and immunisation with live vaccines is liable to give rise to fatal generalised infections. The condition is fatal: in some instances life has been prolonged by transplantation of thymic tissue, but the problem here is to prevent rejection of the grafted thymus by the host T lymphocytes which generate under its influence.

(3) **Combined immunological deficiency.** In **alymphocytic agammaglobulinaemia**, sometimes termed the **Swiss type of agammaglobulinaemia**, both the thymus-dependent and -independent immunity systems fail to develop. The thymus is hypoplastic and deficient in Hassall's corpuscles and small lymphocytes, the lymph nodes are extremely small and lacking in germinal centres, lymphocytes and plasma cells, and circulating lymphocytes are scanty. There

is a near-absence of the three main classes of immunoglobulins from the serum, and both antibody production and cell-mediated immunity are grossly defective. The condition is transmitted as an autosomal recessive character, and affected infants show retarded growth, recurrent bacterial and virus infections, and response to antibiotics and chemotherapy is poor. Immunisation with living viruses is likely to prove fatal, and the condition usually results in death during the first or second year. The basic defect appears to lie in the haemopoietic stem cells (p. 118), which fail to undergo lymphopoiesis.

Combined immunological deficiency occurs also in **reticular dysgenesis**, in which there is a deficiency of haemopoietic stem cells, resulting in failure of lymphopoiesis and haemopoiesis: death usually occurs before or shortly after birth.

In both these conditions, the deficiencies are restored by infusion of haemopoietic cells, which include stem cells, but unless the donor is an HLA-identical sibling there is a grave risk of fatal graft-versus-host reaction (p. 167).

(4) **Other primary immunological deficiencies** are mostly of obscure nature. In **ataxia telangiectasia** there are widespread vascular defects resulting in dilatation of small vessels (telangiectases), and an insidiously-developing immunodeficiency with depression of cell-mediated immunity and low levels of IgE and IgA in the blood. The IgG level is also low in some cases. Recurrent infections of the paranasal sinuses and lungs are the most common consequences of the immunological defect. In some instances the thymus has been found to be poorly developed and lacking in Hassall's corpuscles. The condition appears to be determined genetically by an autosomal recessive gene.

Another condition in which immunodeficiency develops insidiously is the **Wiskott-Aldrich syndrome** in which the platelets are abnormal or reduced in number. There is progressive depletion of lymphocytes in the blood and in the T-dependent areas of the lymphoid tissues. The blood levels of IgM and IgA gradually fall and cell-mediated immunity declines. The condition is determined by a sex-linked genetic defect and affects boys: atopic eczema, attacks of diarrhoea and recurrent infections are common features. Recent reports suggest that administration of Lawrence's transfer factor has a restorative effect on the immunodeficiencies in this condition, in which the thymus appears normal or is slightly diminished in size.

Secondary immunological deficiencies

These are conditions in which the immunity system develops and functions normally but becomes defective from the direct or indirect effect of various disease processes or immunosuppressive agents. Causal conditions include malnutrition, certain infections, various forms of cancer and renal failure.

Susceptibility to infections is a well-known feature of malnutrition, but it is only recently that **protein deficiency**, both experimental and in man, has been demonstrated to impair cell-mediated immune responsiveness. *Because of its prevalence in many parts of the world, this is probably the most important cause of immunodeficiency.*

Depression of cell-mediated immunity may be a feature of various **acute virus infections**, but has been demonstrated most clearly in measles and infectious mononucleosis, in both of which a temporary depression of cell-mediated immunity has been shown by skin tests (e.g. to tuberculin) becoming negative, and by impaired responsiveness of lymphocytes to stimulation *in vitro* by antigens or phytohemagglutinins (see below).

Impaired cell-mediated immunity occurs also in some **bacterial and protozoal infections** in which there is extensive colonisation of the macrophage system, e.g. lepromatous leprosy and leishmaniasis. T-cell function is depressed also in **sarcoidosis**, a condition of unknown cause characterised by tubercle-like granulomas of the lymphoid and various other tissues.

Patients with **advanced cancer** commonly have depression of both T and B cell function: without doubt, this is a result of cancer, although there is evidence that the incidence of cancer (of both the lymphoid and epithelial tissues) is increased in patients who survive with primary immunodeficiencies and in patients on long-term immunosuppressive therapy, e.g. following renal transplantation.

Immunodeficiencies are particularly common in patients with **lymphoid neoplasia (lymphoma)**. In chronic lymphocytic leukaemia, there is very often deficient T and B cell function; this may be due to crowding of the lymphoid tissues, marrow and blood with neoplastic (usually B) lymphocytes. It is, however, of interest that immunosuppression is an early effect of infec-

tion with the reoviruses, which induce lymphomas in animals, and it is likely that human chronic lymphocytic leukaemia (and some other lymphomas) are also virus-induced. Depression of antibody levels is a feature of multiple myeloma, a plasma-cell tumour usually confined to the bone marrow; the high levels of Ig secreted by the myeloma cells increase the rate of Ig catabolism and may also depress antibody responses.

In a third lymphoid neoplasm, Hodgkin's disease, the lymphoid tissues are often extensively infiltrated, and T-cell deficiency is then the usual result: tuberculosis or virus infection (e.g. varicella zoster) may prove fatal.

The immunodeficiency of renal failure affects T-cell, and probably also B-cell, function. This is important in renal transplantation because it helps initially to prevent rejection of the transplanted kidney.

Assessment of immunological function

In cases of suspected immunodeficiency, information can be obtained from examination of the blood to determine: (a) the levels of the various classes of Ig; (b) the presence and titres of ABO blood group antibodies; and (c) the proportions and numbers of T and B lymphocytes.

The responsiveness of lymphocytes to stimulation by antigens, e.g. tuberculo-protein, and to phytomitogens, gives some indication of function. Blast-cell transformation occurs when normal blood lymphocytes are cultured in the presence of phytohaemagglutinin (PHA) or concanavalin A (con-A), both of which stimulate T cells, pokeweed mitogen (PWM) which stimulates both T and B cells, and bacterial endotoxin, which stimulates B cells.

Other tests include assay of antibodies against commonly encountered antigens, and cell-mediated immunity may be investigated by skin tests or *in vitro* techniques (p. 113). Finally, antigens may be administered and the responses measured, but live vaccines should not be used for this purpose in subjects who may not be able to eliminate even attenuated micro-organisms.

With increasing use of immunosuppressive and cytotoxic drugs—cortisone, azathioprene, cyclophosphamide, etc., and also radiotherapy, infections due to immunodeficiencies are becoming common, and often limiting factors in renal transplantation and the treatment of various forms of cancer and other fatal diseases. Some of these agents destroy not only lymphocytes, but also polymorphs and macrophages, and thus depress resistance to infection in more than one way.

References

- Butcher, B. T., Salvaggio, J. E. and Leslie, G. A. (1975). Secretory and humoral immunogenic response of atopic and non-atopic individuals to intra-nasally administered antigen. *Clinical Allergy* 1, 33.
- Cochrane, C. G. and Koffler, D. (1973). Immune complex disease in experimental animals and man. *Advances in Immunology*, Vol. 16, pp. 186–264. Academic Press, New York and London.
- Gibson, T. and Medawar, P. B. (1943). The fate of skin homographs in man. *Journal of Anatomy*, 77, 299–310.
- Ishizaka, K., Ishizaka, T. and Hombrook, M. M. (1966). Physico-chemical properties of reaginic antibody. V. Correlation of reagent activity with γ E-globulin antibody. *Journal of Immunology* 97, 840–53.
- Ishizaka, T., Ishizaka, K. and Tomioka, H. (1972). Release of histamine and slow reacting substance of anaphylaxis (SRS-A) by IgE—anti-IgE reactions on monkey mast cells. *Journal of Immunology* 108, 513–20.
- Larner, J. L. (1977). *Cyclic Nucleotide Metabolism*, pp. 52. In *Current Contents Series*, Upjohn Co., Kalamazoo, Michigan.
- Matthew, D. J., Norman, A. P., Taylor, B., Turner, M. W. and Soothill, J. F. (1977). Prevention of eczema. *Lancet* i, 111–13.
- Reid, F. M., Sandilands, G. P., Gray, K. G. and Anderson, J. R. (1979). Lymphocyte emperipolesis revisited. *Immunology* 36, 367–72.
- Taylor, B., Norman, A. P., Orgel, H. A., Turner, M. W., Stokes, C. R. and Soothill, J. F. (1973). Transient IgA deficiency and infantile atopy. *Lancet* i, 111–13.
- Wilkinson, P. C., Parrott, D. M. V., Russell, R. J. and Sless, F. (1977). Antigen-induced locomotor responses in lymphocytes. *Journal of Experimental Medicine* 145, 1158–68.

Further Reading

- Gell, P. G. H., Coombs, R. R. A. and Lachmann, P. J. (Eds.) (1975). *Clinical Aspects of Immunology*, 3rd edn., pp. 1754. Blackwell Scientific, Oxford. (Extensive reviews by leading workers.)
- Holborow, E. J. and Reeves, W. G. (Eds.) (1978). *Immunology in Medicine*, pp. 1185. Academic Press, London; Grune and Stratton, New York. (A comprehensive but readable account of disease processes with an immunological basis.)
- Irvine, James (Ed.) (1979). *Medical Immunology*, pp. 506. Teviot Scientific Publications, Edinburgh. (An up-to-date readable account by leading workers.)
- Miescher, P. A. and Müller-Eberhard, H. J. (Eds.) (1976). *Textbook of Immunopathology*, 2nd edn., pp. 1118. Grune and Stratton, New York, San Francisco and London. (Extensive reviews by leading workers.)
- Rose, N. R., Milgrom, F. and van Oss, C. J. (Eds.) (1978). *Principles of Immunology*, 2nd edn., pp. 544. Macmillan Publishing Co. Inc., New York. (An up-to-date account by leading workers.)
- The HLA System* (1978). *British Medical Bulletin* **34**, pp. 213–316. (A series of review articles on biological and clinical aspects of the system.)
- See also* Bibliography for Chapter 5 (pp. 139–40).

Host—Parasite Relationships

Throughout evolutionary development, many species have adapted to a parasitic existence, living in or on the surface of a host of another species, from which they derive warmth, nourishment and mobility. The relationship is not necessarily harmful to the host, and may be advantageous. For example, various relatively harmless bacteria colonise the skin of man and help to exclude more harmful bacteria, while reabsorption of bile pigment from the gut and the production of vitamin K depend largely on the metabolic activities of the intestinal bacterial flora. These normal inhabitants of the skin and mucous membranes are called **commensals**. Other parasites, termed **pathogens**, are less well adapted and by injuring the host endanger their own survival: they include many species of micro-organisms (microbes) including viruses, bacteria, fungi and protozoa and also metazoa of various sizes. The terms **pathogenicity** and **virulence** are commonly used synonymously to indicate the capacity of a particular micro-organism to cause disease.

Although it is important to distinguish between commensals and pathogens, the distinction is not absolute, for many commensals are only harmless so long as they are kept at bay by the host's defence mechanisms. In immunodeficiency states, for example, various normally harmless microbes may behave as 'opportunistic' pathogens. Similarly, a breach of local defence mechanisms, even in a normal individual, may allow commensals to cause severe infections, an example being *Escherichia coli*, which normally inhabits the gut: this bacterium may be introduced into the urinary tract by catheterisation of the bladder, and may then cause severe acute pyogenic inflammation, even extending into the kidneys. Local abnormalities in the host may also predispose to injury by commensals: for instance, heart valves which have been scarred and distorted by rheumatic

fever are readily colonised by *Streptococcus viridans*, a bacterium which lives in the mouth and finds its way into the blood following tooth extraction, or even when the teeth are brushed vigorously. In normal individuals it is quickly eliminated, but it can settle and multiply in the distorted valve cusps, causing bacterial endocarditis. Because the distinction between pathogens and commensals is not sharp, it is helpful to use the term **infection** to indicate the presence of a particular type of micro-organism in a part of the body where it is normally absent, and where, if allowed to multiply, it is likely to be harmful, i.e. to cause **infective disease**.

As implied above, most infective diseases depend on penetration of the host's tissues by micro-organisms, and the factors concerned in such invasion provide the first major topic of this chapter. Following invasion, the microbes may be eliminated without causing obvious disease (inapparent infection) or clinical disease of any grade of severity may follow: the factors determining these events form a second major topic. Lastly, two important reactions to infection, neutrophil leukocytosis and fever, will be considered.

The subject of infective disease is extremely complex, involving as it does a consideration of the relationships between man and numerous species and strains of micro-organisms. The following account is limited to a brief outline of the subject.

Factors determining invasion

The skin and mucous membranes are exposed to many different types of micro-organisms present in expired droplets in the air, in dust particles, and in food and water. The skin and various mucous membranes on which these organisms settle have properties which render

them suitable for the survival and sometimes multiplication of certain organisms, but inhospitable to others. In some instances, the requirements of a particular microbe for growth *in vitro* help to explain its colonisation of particular parts of the surface of the body, but many of the factors determining such colonisation are still unknown, and indeed the predilection of certain bacteria for a particular host species is in most instances quite unexplained. Nevertheless, certain factors are known to be of great importance in limiting or preventing invasion by many types of microbes, and these must be considered briefly.

Barriers to invasion

(a) **Mechanical barriers.** The superficial keratinised layer of the epidermis is an excellent mechanical barrier to microbial invasion, and provided it is kept clean and dry, direct invasion is extremely unlikely. Penetration may however occur when dirt is allowed to accumulate on the skin and particularly in moist warm areas subject to friction, such as the axillae and sub-mammary folds. In many skin diseases which result in exudation with loss or sogginess of the keratin layer, bacterial and fungal infections are common complications. The conjunctival, oral, respiratory-tract and gastro-intestinal mucosae, covered as they are by a film of mucous or serous secretion, also present a formidable barrier to many micro-organisms, although some can penetrate the epithelium readily, e.g. influenza virus, rhinoviruses.

Wounds and ulcers of the skin and mucous membranes open up pathways for bacterial invasion and are obviously important causes of infection. Burns are particularly liable to become heavily infected because the dead superficial tissue provides a good medium for coliform bacilli, staphylococci, pyocyanea and many other bacteria. In the mouth, tooth extraction and tonsillectomy inevitably lead to bacterial invasion, and tonsillectomy has been shown to predispose to invasion by the virus of poliomyelitis in the post-operative period. Vitamin A and C deficiencies also impair the resistance of the mucous membranes and skin to bacterial invasion.

Some parasitic organisms have evolved a life cycle in which they multiply in insect vectors

and are introduced to man and other hosts by the insect bite, thus penetrating the major barrier of the skin. Examples include the protozoa which cause malaria, the metazoan filarial worms, and the virus of yellow fever, all of which are transmitted by mosquitoes. *Yersinia pestis*, the cause of bubonic plague (the Black Death), is transmitted by the flea of the black rat, and the rickettsiae which cause typhus are spread by ticks, mites and lice. Rabies virus enters the tissues by the bite of a rabid animal.

(b) **Glandular secretions.** The secretions of glands opening on to the skin surface play an important role by maintaining the integrity of the skin, and also by providing an environment in which many types of bacteria cannot survive for long. The acidity of the sweat and the long-chain unsaturated fatty acids produced by the action of commensal bacteria on sebaceous secretion both exert a selective bactericidal effect, and consequently the bacterial flora of the skin surface tends to be rather constant: it has been shown that some types of pathogenic bacteria, when placed on the skin, are virtually all destroyed within an hour or two. The secretions of mucous membranes possess similar qualities. **Lysozyme**, an enzyme which digests the mucopeptide of bacterial cell walls, is present in high concentration in the lacrimal gland secretion and probably exerts an important protective effect in the conjunctival sac: it is secreted also by the salivary and nasal glands but in much smaller amounts. **Antibodies of IgA class**, modified by addition of a 'transport piece' so that they are resistant to digestive enzymes, are present in saliva, tears, intestinal contents, respiratory tract mucus, milk and urine (p. 109). Provided that IgA antibody has developed against a particular organism as a result of previous infection, it will be represented in these secretions. This is of importance in preventing invasion by certain viruses, for the virus may encounter the antibody in the surface mucus and be neutralised by it: its significance in relation to bacterial invasion is less certain, although there is evidence that IgA antibody may render bacteria highly susceptible to the lytic action of lysozyme, and it also activates complement by the alternative pathway.

The acidity of the gastric juice is effective in killing most types of microbes ingested in food or water; but hypochlorhydria due to chronic gastritis is common, and minor illnesses and

even emotional stresses can reduce temporarily the acidity of the juice. In general, those microbes which cause intestinal infections, such as the salmonellae and dysentery bacilli, are relatively acid-resistant. *Entamoeba histolytica*, the cause of amoebic dysentery, produces cysts which resist the gastric juice and pass through the stomach before hatching out and invading the wall of the colon.

The normal acidity of the urine contributes to the defences of the urinary tract against infection. Also there is a mucosal factor which eliminates bacteria in contact with the urinary tract epithelium but its nature and mechanism are not understood.

(c) Secretion currents. The continuous flow of tears over the the surface of the conjunctiva has an important effect in the removal of contaminating bacteria, which are carried rapidly into the nasopharynx. In the nose and mouth also, the secretions covering the mucosa flow towards the pharynx and hence to the stomach, carrying with them residual food particles, bacteria, etc. The importance of the saliva is illustrated by the oral infections and severe dental caries which accompany loss of salivary secretion in Sjögren's syndrome (p. 594). The lacrimal secretion is also diminished, and conjunctival infections result. The importance of removal of contaminating bacteria by the saliva may explain the common occurrence of infection in the crypts of the tonsils and also in the periodontal sulci, for once bacteria gain entrance to these spaces, they are out of the main stream of salivary flow.

In the respiratory tract there is a continuous flow of mucus upwards over the surface of the bronchial and tracheal mucosa: inhaled particles are caught up and removed in this stream, and the air is almost sterile by the time it reaches the respiratory bronchioles. This defence mechanism is dependent on a normal production of mucous secretion and on the integrity of the ciliated respiratory epithelium. Most of the micro-organisms capable of invading the respiratory mucosa in spite of mucociliary flow are enabled to do so by having surface components which allow them to bind to respiratory epithelium: such organisms include influenza virus, *Mycoplasma pneumoniae*, rhinoviruses (common cold) and *Bordetella pertussis* (whooping cough). Other micro-organisms are less likely to cause respiratory in-

fections unless the mucosa is first damaged. Such damage may be caused by the virus of influenza which parasitises the respiratory epithelium, interfering with its protective function: as a result, secondary bacterial infection invariably develops, and by extending into the alveoli may give rise to pneumonia. The integrity of the respiratory mucosa is also seriously impaired in chronic bronchitis, most commonly due to cigarette smoking but also to atmospheric pollution: this leads to metaplasia, the ciliated epithelium being replaced by goblet cells in the smaller bronchi. There is increase in the amount of secretion, which also becomes more viscous, and this tends to stagnate and become infected.

Intestinal pathogens, such as the salmonellae of 'food poisoning' and the shigellae of bacillary dysentery, induce an acute inflammatory reaction in the intestinal mucosa: diarrhoea results from the increased peristalsis and exudation, and repeated evacuation of the gut helps to get rid of the offending bacteria.

The flow of urine is of importance in preventing growth and spread of any bacteria gaining entrance to the urinary tract by the urethra, and any abnormality resulting in stagnation of urine or incomplete emptying of the bladder, particularly if chronic, e.g. obstruction by an enlarged prostate, predisposes to infection.

(d) Bacterial commensals. In spite of the defence mechanisms described above, the skin, mouth, nasal cavity, conjunctival sac and intestines are all colonised by bacteria of various types. The local environment provided by each of these various surfaces favours the survival of particular types of bacteria and thus each regional surface develops its own flora. In their usual site of colonisation, most of these commensals are non-pathogenic, and they tend to prevent the establishment of other types of microbes, including pathogens, by competing for nutrients and by release of metabolic products which are toxic to other organisms.

In normal circumstances the bacterial floras of the various surfaces are remarkably stable, but if they are disturbed, colonisation by pathogens may result: hence the common occurrence of fungal infections of the pharynx in patients on antibiotic therapy, and the production of lesions by the toxin of *Clostridium difficile* in pseudomembranous colitis which

may arise when the normal flora is depressed by broad-spectrum antibiotics. 'Seeding' of the gut with non-pathogenic bacteria has achieved some success in preventing the overgrowth of pathogens in neonates and in patients treated by antibiotics.

(e) **Phagocytes.** There is evidence that phagocytic cells migrate on to the surface of various mucous membranes: for example neutrophil polymorphs pass through the thin epithelium lining the depths of the tonsillar crypts, and macrophages pass into the alveoli of the lungs. In both these sites the migrant cells have been shown to phagocytose particles on the surface of the mucosa and this may play a role in preventing invasion.

Invasive capacity of micro-organisms

Micro-organisms vary greatly in their capacity to invade the host's defensive barriers. Most bacteria cause injury only after invading the host's tissues, but some are virtually incapable of invasion and yet can produce disease. For example, *Clostridium tetani*, the cause of tetanus, flourishes only in dead tissue, foreign material and exudate in wounds, but its toxin is absorbed and has serious effects on the nervous system. *Vibrio cholerae* does not invade the mucous membrane of the small intestine, but secretes a toxin which, by disturbing the control of fluid transport across the epithelium, causes severe dehydration. Other organisms, and particularly some viruses, are very highly invasive and infect virtually all individuals who have not previously encountered or been immunised against them, e.g. the viruses of morbilli (measles) and rubella (german measles). Examples of highly invasive bacteria include *Yersinia pestis* (the cause of plague), *Salmonella typhi* (typhoid fever) and the brucellae (undulant fever), all of which regularly invade the bloodstream. However, a great many bacteria lie intermediate between these extremes in their invasive capacity. This includes the more important pyogenic bacteria which are commonly present in the nose or throat, or on the skin. Their presence is often harmless, but disturbances of defence mechanisms may allow them to invade and cause lesions.

In general, bacteria of high invasive capacity are also highly pathogenic, but there is little correlation between the invasive capacity and

pathogenicity of viruses. For example, poliovirus invades readily but only a small proportion of infected individuals develop clinical disease, and non-pathogenic strains are administered orally to produce infection and immunity. Also the protozoon *Toxoplasma gondii* is highly invasive and yet, apart from the lesions it causes in fetal life, it is of low pathogenicity.

Pathogenic effects of micro-organisms

Bacteria which have invaded the host tissues may be destroyed without causing clinically apparent disease, may promote a local inflammatory lesion, or may spread to other parts of the body and produce widespread lesions. The two major ways in which bacteria are known to cause pathological changes are firstly by the production of toxins, and secondly by promoting hypersensitivity reactions on the part of the host.

Viruses cause injury by invading the host's cells and utilising the cellular synthetic processes for their own replication. The colonised cells may be destroyed directly by the replicating virus, or as a result of an immunological host reaction (probably mainly delayed hypersensitivity).

Bacterial toxins

These are of two main types, exotoxins and endotoxins.

Exotoxins are secreted by living bacteria: they are simple proteins, are often extremely potent, and vary considerably in their biological effects upon the host. They are antigenically specific and their biological activity is usually neutralised by union with antibody. Many pathogenic bacteria produce a number of different exotoxins when cultured *in vitro*. Thus *Streptococcus pyogenes* and *Staphylococcus aureus*, two of the most important pyogenic bacteria, produce haemolysins and hyaluronidases. *Strep. pyogenes* also produces a leukocidin which kills leukocytes, and *Staph. aureus* a coagulase which clots fibrinogen. Some exotoxins are injurious to virtually all types of host cell and their effects thus depend on their concentration and distribution. *Corynebacterium diphtheriae*, the cause of diphtheria, secretes such a toxin and at the site of infection, usually

the pharynx, it causes local tissue necrosis. Less florid but still severe cell injury is far more widespread and is reflected morphologically in fatty change and necrosis of the parenchymal cells of the various organs: in severe cases, death may result from its effect upon the myocardium (Fig. 7.1). The mechanism of injury by this particular toxin is known (p. 8): other toxins with a similarly widespread effect are produced by many of the pathogenic Gram + ve bacteria but in most instances the mechanism of toxic action is not known. Some have enzymic activity, e.g. phosphatases, proteases, lipases. Some bacteria produce toxins which act specifically on one type of tissue, e.g. the neurotoxins of *Cl. botulinum* interfere with the production of acetylcholine at cholinergic synapses in the peripheral nervous system, and cause a flaccid paralysis, while the neurotoxin of *Cl. tetani* has a contrasting effect on the synapses in the central nervous system, resulting in widespread tetanic muscular contractions in response to slight local stimuli.

Attempts to equate the pathogenic effects of

a particular micro-organism with its toxins have encountered difficulties: not only are many toxins produced by a single strain of bacteria but different samples of a toxin, even in highly purified form, may have different biological properties. Also toxins vary greatly in their effects on hosts of different species, and experimental observations are not necessarily applicable to man. Finally, production or non-production of toxin by bacteria growing *in vitro* does not necessarily indicate a similar behaviour *in vivo*. It is a feature of exotoxins that their biological effects are neutralised by the corresponding antitoxin, and in some instances, e.g. diphtheria and tetanus, prior administration of the antitoxin or active immunisation by injection of *toxoid* (inactivated toxin which maintains its immunogenicity) will protect animals against the effects of injection of the toxin and man against the disease. Thus in some instances, particular toxins have been incriminated beyond all reasonable doubt as the pathogenic agents responsible for the disease; in others, it seems most likely that toxins are responsible, but there remains the possibility that the bacteria may have other, at present unknown, pathogenic properties in addition to toxin production.

Endotoxins are structural elements of bacteria and are released only when the bacterium dies. They are constituents of the cell wall of Gram-negative bacteria, and are complexes of phospholipid (lipid A), polysaccharide and protein. The endotoxins produced by different Gram-negative bacteria are antigenically different but they all have the same biological effects and the active component resides in lipid A. Endotoxin is responsible for fever, activation of complement by the alternative pathway (p. 143), intravascular conversion of fibrinogen to fibrin, vascular lesions and cellular necrosis in various organs. In small dosage it causes a neutrophil leukocytosis, in large amounts a leukopenia followed by a leukocytosis. In severe Gram-negative bacterial infections, a state of shock develops with some or all of the above features—'septic' or 'endotoxic' shock (p. 264). Because they produce fever, endotoxins are sometimes termed *pyrogens*. They are heat-stable and unless special precautions are taken are liable to contaminate apparatus used for haemodialysis, etc., and fluids prepared for parenteral administration.



Fig. 7.1 Heart muscle in fatal diphtheria, showing destruction and disappearance of muscle fibres and a light inflammatory cellular infiltrate. $\times 115$.

Hypersensitivity reactions to micro-organisms

Virtually all microbial infections stimulate immune responses by the host, and the reaction of the antibodies or primed T lymphocytes with microbial antigens can result in hypersensitivity of various types. Atopic (type 1) reactions, such as urticaria, are a common feature of infestation by parasitic worms, even in individuals not otherwise predisposed to atopy, and microbial infections sometimes cause atopic reactions in individuals predisposed to this type of hypersensitivity. Cytotoxic antibody (type 2) reactions may, in theory, result from the cross-reaction of microbial-induced antibodies with host cells, a possible example being rheumatic fever, in which antibodies to *Strep. pyogenes* react with heart muscle cells. Immune complex (type 3) reactions are important complications of some infections. Local Arthus reactions occur when microbial antigens in infected tissues react with antibodies in the plasma. Immune complexes are deposited in the walls of small blood vessels. Circulating immune-complex disease occurs when microbial antigens enter the blood, immune complexes then being formed in the plasma, resulting in an acute febrile reaction and/or deposition of the complexes in small blood vessels, notably in the glomeruli where they are responsible for glomerulonephritis (p. 156). The acute generalised reaction is seen in the severe form of dengue fever which occurs in people who have had a previous, usually mild infection with the virus. On re-infection (possibly with another strain of the virus) large amounts of viral antigen encounter antibody in the plasma, triggering off the complement, clotting, plasmin and kinin systems (p. 55) and causing a profound state of shock. Other examples of infections giving rise to immune-complex reactions, and the way in which lesions are produced, are described on pp. 154–6.

Cell-mediated immunity is an important defence mechanism in various infections. The lymphokines released when primed T cells react with microbial antigens (p. 158) are responsible for both destruction of micro-organisms by macrophages and the tissue injury of delayed (type 4) hypersensitivity. Consequently, the two phenomena are commonly associated. The classical example is tuberculosis, in which cell-mediated immunity develops to tuber-

culoprotein, and is largely responsible for the lesions of this disease. *Myco. tuberculosis* has not been shown to produce toxins and can colonise macrophages in culture without causing apparent injury: addition of primed T lymphocytes reactive with tuberculo-protein results in destruction of macrophages and their ingested micro-organisms. The morphological features of tuberculosis, described in the next chapter, can all be explained on the basis of delayed hypersensitivity. Leprosy is another disease in which delayed hypersensitivity plays a major role. In some cases, cell-mediated immunity is weak or absent and *Myco. leprae* multiply progressively, mostly within macrophages. Like tubercle bacilli, they cause little or no cell injury, and the lesions consist of enlarging nodules composed of macrophages containing large numbers of *Myco. leprae*. In other cases, strong cell-mediated immunity develops, and the bacteria are kept partly in check. Very few are demonstrable in the lesions, but delayed hypersensitivity results in tissue necrosis and fibrosis (p. 215). This is a good example of the dual effect of cell-mediated immunity—it limits the numbers of micro-organisms, but also causes injury of host tissues.

Delayed hypersensitivity reactions are commonly prominent in fungal, viral and chronic bacterial infections. They are probably involved also in some of the skin lesions of chickenpox, measles, etc. but firm evidence on the pathogenesis of the skin rashes in these conditions is remarkably scanty.

Defence mechanisms in infections

When micro-organisms have invaded the tissues, there are three major defensive reactions which tend to limit their multiplication and spread, and bring about their destruction: these are the inflammatory reaction, phagocytic activity and specific immune reactions.

The inflammatory reaction

The acute inflammatory reaction. The defensive role of this reaction has been considered in Chapter 3. Without doubt, it is of considerable importance, and *those infections which are accompanied by acute inflammation at the site of invasion are more likely to remain localised than those in which invasion is accomplished without*

local injury or reaction. The pyogenic bacteria are a common cause of the former type of infection, while silent invasion is illustrated by *Treponema pallidum*, the cause of syphilis, which spreads widely through the body before the appearance of a local lesion at the site of entry. Other bacteria which may enter the body silently and spread widely include *Neisseria meningitidis*, a cause of acute meningitis, and brucellae, the cause of undulant fever. Many of the parasites transmitted by biting insects, such as the plasmodia which cause malaria, produce generalised infection without a significant local reaction, and many viruses invade the body and produce viraemia without first producing local inflammation.

Chronic inflammatory change also plays a defensive role by exposing the micro-organisms to phagocytes, antibodies and effector T lymphocytes, and by surrounding them with a layer of granulation tissue which has been shown to be an effective barrier to bacteria. In the more prolonged infections, such as tuberculosis, surviving micro-organisms may be effectively confined within a zone of dense fibrosis resulting from chronic inflammatory change.

Phagocytosis and killing of micro-organisms

A general account of phagocytosis has been given on p. 63, and we are concerned here with

factors which determine the capacity of neutrophil polymorphs and macrophages to phagocytose and subsequently kill micro-organisms. Bacteria differ greatly in their resistance to these processes, and such resistance is often an important factor in their pathogenicity.

The inflammatory and immune responses are important host factors favouring phagocytosis and killing of micro-organisms. They help to provide an environment favourable to phagocytosis, render the micro-organism more susceptible to phagocytosis, and increase the phagocytic and killing activities of phagocytes. These factors are illustrated for macrophages in Fig. 7.2. Emigration of polymorphs and monocytes is part of the inflammatory reaction, and the inflammatory exudate opens up tissue spaces in which the emigrated phagocytes can move. Immunoglobulins and components of complement enter infected tissues in the inflammatory exudate; if specific antibodies are present, they may aid phagocytosis by rendering the micro-organisms more susceptible to phagocytic ingestion (opsonisation), or by neutralising toxins harmful to phagocytes. The reaction of antibodies usually activates complement, which may kill micro-organisms directly, or favour their destruction by enhancing the inflammatory reaction and chemotaxis of phagocytes (pp. 55, 60).

Both polymorphs and macrophages have

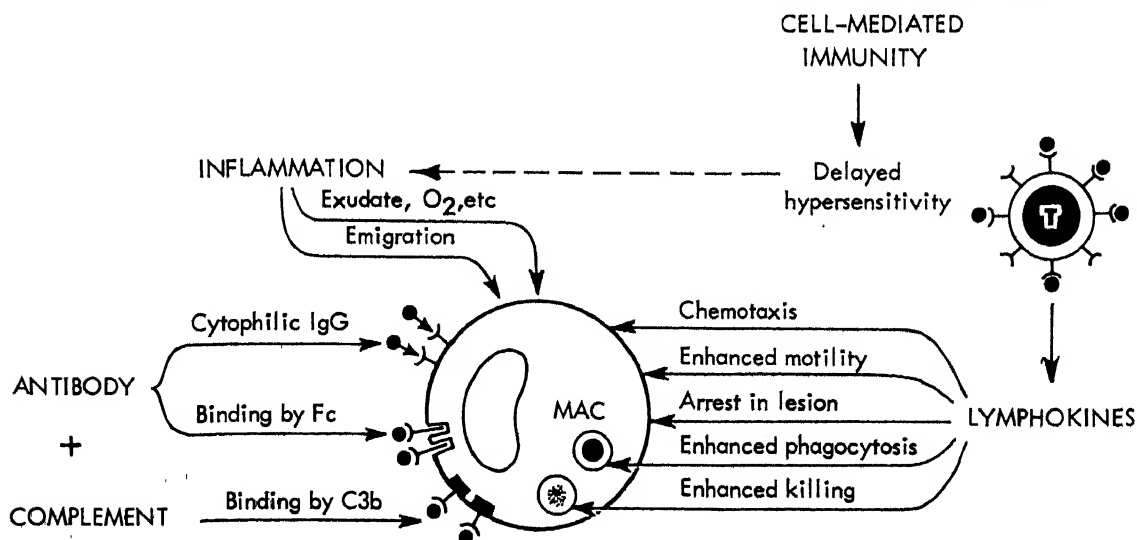


Fig. 7.2 The ways in which the inflammatory reaction, delayed hypersensitivity and antibodies promote the emigration and aggregation of macrophages at the site of infection and enhance their binding, phagocytic and killing capacities for micro-organisms (represented by black spheres). Y represents specific receptors for microbial antigens on T or B lymphocytes and also antibodies binding by various receptors to the surface of macrophages.

surface receptors for the Fc component of IgG antibodies, and this facilitates surface binding and subsequent phagocytosis of microbes sensitised with IgG antibodies (Fig. 7.2): they also have surface receptors for the C3b component of reacted complement but, surprisingly, while complement fixation by antibody-sensitised bacteria enhances binding to the phagocyte, it apparently does not enhance phagocytosis. IgM antibodies are opsonic, particularly for micro-organisms with a non-protein capsule, but their opsonic effect is not due to specific binding to phagocytes, which do not have surface receptors for Fc of IgM.

Cell-mediated immunity is particularly effective in destroying microbes which invade host cells. The lymphokines released when primed T cells react with antigen include factors which promote inflammation, accumulation of macrophages and lymphocytes, and enhance, both specifically and non-specifically, the phagocytosis and killing of ingested micro-organisms by macrophages (see below).

The mechanisms of killing of micro-organisms by phagocytes are complex. Unlike phagocytosis, which occurs readily under anaerobic conditions, killing is accompanied by increased oxygen uptake. In **polymorphs**, oxygen is converted by NADPH into superoxide ($^-\text{O}_2$) by removal of an electron; some of this is converted into hydrogen peroxide (H_2O_2) and singlet oxygen ($^1\text{O}_2$), which has an unstable distribution of electrons around the two nuclei, is also produced. All these forms of highly reactive oxygen are produced within the phagosome, the membrane of which protects the host cell from their effects. They react with the wall of the phagocytosed micro-organism and are highly lethal to many bacteria, viruses and fungi. Hydrogen peroxide also co-operates with myeloperoxidase which, together with halogen ions, forms a system which attacks the microbial cell wall.

In addition to the above mechanisms, the low pH within phagosomes is unfavourable to many micro-organisms, and other lysosomal products exert a harmful effect, notably cationic lysosomal proteins which injure microbial cell walls, lactoferrin (an iron-binding protein), and lysozyme which has a synergistic lytic effect with complement.

Macrophages lack myeloperoxidase, cationic microbicidal proteins and lactoferrin. They are

nevertheless capable of producing microbicidal forms of oxygen and they are activated by bacterial products and by the lymphokines produced by T lymphocytes in delayed hypersensitivity reactions (p. 158); their motility, phagocytic activity, lysosomal enzymes and killing capacity are all increased as a result.

Microbial resistance to phagocytes. In general, those bacteria which develop a non-protein capsule, e.g. the anthrax bacillus or smooth strains of *Strep. pneumoniae* and *Haemophilus influenzae*, are not readily phagocytosed. Some bacterial products are chemotactic, but *Strep. pyogenes* and some other bacteria secrete toxins which injure phagocytic (and other) cells and so inhibit phagocytosis. Other bacterial products, e.g. the endotoxins of Gram-negative bacteria, enhance the phagocytic activity of polymorphs and macrophages in low concentrations, but inhibit it in higher concentrations. *Staphylococcus aureus* secretes a factor ('Protein A') which partially inhibits phagocytosis of bacteria sensitised by antibody, possibly by blocking their binding to the Fc receptors of phagocytes (see above).

A number of micro-organisms undergo phagocytosis but are able to resist the microbicidal activity of phagocytes and even multiply within them. A good example is provided by the gonococcus (the cause of gonorrhoea) which colonises neutrophil polymorphs (Fig. 7.3). Some

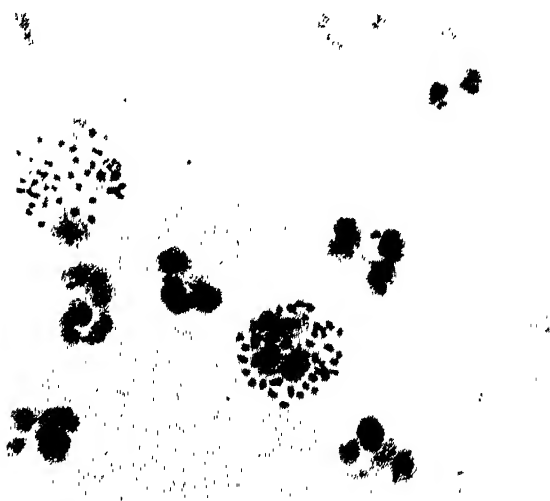


Fig. 7.3 Smear of urethral exudate in acute gonorrhoea. Two polymorphs contain large numbers of gonococci, and show degenerative changes. Other polymorphs contain few or no bacteria and appear relatively healthy. (Gram stain.) $\times 1200$.

viruses undergo phagocytosis by macrophages, but can bind to the phagosomal wall and pass into the cytosol of the host cell. A number of organisms succeed in preventing the fusion of lysosomes with the phagocytic vacuole, and so protect themselves from lysosomal microbicidal products: this is observed with tubercle bacilli ingested by macrophages in culture and with the protozoon *Toxoplasma gondii* and the fungus *Aspergillus flavus*, all of which cause chronic infections in man. In other instances, the organism can flourish within phagosomes, e.g. the brucellae which cause undulant fever. Other organisms have a suppressive effect on cell-mediated immunity and persist within macrophages which are handicapped by lack of the enhancing effects of T-cell lymphokines on their microbicidal activity. This includes the leprosy bacillus and some protozoa.

The immune response

The several ways in which antibodies and cell-mediated immunity help to destroy micro-organisms have been described in this and preceding chapters, and may be summarised as follows. **Antibodies** of IgA class are important in preventing the invasion of mucous membranes by viruses and probably by some bacteria (p. 175). IgM and IgG antibodies can neutralise bacterial toxins, agglutinate and immobilise micro-organisms (p. 108), and prevent cell invasion by viruses: by activating complement, they may cause lysis of microbial cell walls without the intervention of phagocytes (p. 144). Activation of complement also promotes the inflammatory reaction and attracts polymorphs by chemotaxis (pp. 55, 60). Antibodies, particularly those of IgG class, also opsonise micro-organisms, thus favouring their ingestion and destruction by phagocytes (see above).

When the primed T lymphocytes produced by **cell-mediated immune responses** react with microbial antigens, they release lymphokines which induce the inflammatory and other changes of the delayed hypersensitivity reaction (p. 158). In addition to exerting chemotactic and immobilising effects on macrophages, lymphokines include a macrophage-activating factor which increases their killing capacity for ingested micro-organisms, and a second factor—the specific macrophage-arming factor

—which enables macrophages to kill allogeneic target cells and may also mediate destruction of micro-organisms.

One of the main purposes of this summary is to emphasise the complex relationships and synergism between the inflammatory response, phagocytosis, and immunological reactions, which together provide a closely interwoven system of defence against micro-organisms.

Interferon (p. 194) is probably mainly responsible for arresting virus infections, yet children with congenital T-cell deficiencies tend to develop progressive virus infections. This could be attributable to loss of the interferon which is produced by T cells responding to antigen, although many other types of cell are capable of producing interferon.

Microbial resistance to the host's immune response. Micro-organisms which colonise host cells are protected from **antibody** in the plasma and tissue fluids, and can persist in spite of a strong antibody response. This is illustrated by the brucellae of undulant fever, the protozoon *Leishmania donovani* which causes leishmaniasis, and some fungi, all of which can survive in macrophages. This mode of protection is particularly successful for organisms which do not kill the host macrophage nor prevent its division. As obligatory intracellular parasites, viruses are protected from antibodies once they have established an infection, although many of them are prevented by antibody from invading the host and spreading by the bloodstream. Those viruses which are integrated into the host cell genome (p. 302) may not provide an antigenic stimulus unless they replicate, and indeed have been shown in various species of animals to be transmitted from generation to generation via the host's germ cells.

To survive, intracellular micro-organisms must also protect themselves against the host's **cell-mediated immune response** and many of them do this by exerting a suppressive effect on cell-mediated immunity in general. This is observed in a number of viral infections, including measles, mumps, infection with the Epstein-Barr virus (the cause of glandular fever), and the animal leukaemia viruses which colonise lymphoid cells. It is also a feature of lepromatous leprosy (p. 216), leishmaniasis (p. 564) and malaria (p. 305).

In some infective diseases, sufficient microbial antigen is produced to overwhelm the

immune response. This is seen in severe acute cases of meningococcal septicaemia and meningitis, in which free bacterial polysaccharide antigen can be detected in the serum and CSF (a test used in diagnosis), and also in acute pneumococcal pneumonia or septicaemia. Another example is provided by the B hepatitis virus (p. 670) which replicates in the liver and produces large amounts of viral envelope: in carriers, free viral surface antigen can be demonstrated in the serum. The cell-mediated immune response can also be overwhelmed in extensive infections, for example in widespread tuberculosis or acute tuberculous bronchopneumonia; in these conditions the skin test with tuberculin, which is based on a delayed hypersensitivity reaction (p. 157), becomes negative.

Another ingenious method of circumventing the host's immune response is by **antigenic variation**. The best-known examples are relapsing fever, caused by *Borrelia recurrentis* and trypanosomiasis caused by flagellate protozoa. Both of these organisms stimulate an antibody response to which they are susceptible, but they possess a number of genes coding for antigenically-distinct surface coat material. Although the great majority are destroyed by the host's antibody response, occasional organisms operate another gene for surface coat production and arise as resistant variants with consequent relapse, and so the diseases are characterised by successive recurrences. The influenza virus is notorious for producing mutant strains which are responsible for fresh outbreaks in the population.

Another elegant method of protection is illustrated by the group of parasitic trematodes termed *Schistosomes*. These parasites enter the body as a larval form or schistosomule, which stimulates an immune response in the host, but the larvae rapidly become coated with host blood-group substances and so provide themselves with an immunological cloak which they maintain during development and adult life. The immune response is partly effective in destroying schistosomules subsequently entering the body, before they become coated, and further infection is thus usually avoided. Unfortunately many of the eggs produced by the adults are arrested in the host tissues and elicit a delayed hypersensitivity reaction resulting in a chronic inflammatory lesion, notably in the liver (p. 698).

Mention should also be made of the possibility of a micro-organism inducing in the host **specific immunological tolerance** to its own antigens. This occurs when mice are infected via the ovum with lymphocytic choriomeningitis virus. The virus induces tolerance and persists into adult life, notably in the ependyma and cells lining the meninges: little or no harmful effects result until tolerance partly breaks down and the development of antibodies is followed by deposition of antigen-antibody complexes in the glomeruli and elsewhere. If a mouse is first infected some time after birth, it develops choriomeningitis due to the development of cell-mediated immunity and a consequent delayed hypersensitivity reaction with viral antigens. A parallel in man has not been demonstrated. Tolerance might arise also from **molecular mimicry**, in which a micro-organism possesses surface antigens closely similar to host antigens. Some of the coliform bacilli in the gut have been shown to share a common antigen with colonic epithelium, and antibodies induced by various bacteria react with particular transplant (HLA) antigens of man. Such antigenic similarities have not been shown to influence infections, but individuals possessing particular HLA antigens have been found to be unduly prone to certain infections. For example, Reiter's syndrome (p. 988) due to a chlamydia, occurs especially in people with HLA-B27, and tolerance or genetically-determined inability to respond to a particular microbial antigen remain as possibilities.

Antibacterial drug therapy

In many types of bacterial infection, antibiotic drugs are capable of killing the bacteria or suppressing their multiplication, and thus tip the balance in favour of the host and terminate the infection. Antibiotic therapy is not, however, without risk. Apart from their toxic side-effects and the induction of hypersensitivity reactions, antibiotics can, as already mentioned, encourage the multiplication of resistant strains of bacteria. Such resistance may be determined by orthodox genetic bacterial inheritance, but one form of resistance can be transmitted from resistant to susceptible bacteria by transfer of *plasmids* (extra-chromosomal genetic factors) from antibiotic-resistant bacteria. Antibiotic therapy has, in some circumstances, resulted in

a great increase in the incidence of antibiotic-resistant infections, particularly among hospital patients, and this emphasises the importance of avoiding their indiscriminate use.

Diminished resistance to infection

There is a wide range of individual 'natural' resistance to microbial infections among apparently normal individuals. This is largely unexplained, but it is apparent that variations in susceptibility to some infections are related to the individual's 'transplant' iso-antigens (see above). In passing, it is worth noting that similar associations with HLA antigens have recently been claimed for many diverse diseases of obscure nature, e.g. disseminated sclerosis, the connective tissue diseases and lymphoid neoplasia. These observations may throw light on genetic factors in disease, and help to explain the nature of 'constitutional' predisposition and resistance.

Impaired resistance to infection can occur in many ways. Examples of defects in local barriers to invasion have been given earlier in this chapter. Normal phagocytic function, immune responsiveness and complement function are all

essential factors contributing to the range of protective mechanisms. A serious fall in the number of polymorphs in the blood, as in agranulocytosis, predisposes to bacterial invasion and severe, spreading infections, e.g. of the pharynx and intestine, by both commensals and more highly pathogenic bacteria. Genetically-determined defects of polymorph function also predispose to bacterial infection. These polymorph abnormalities are considered in Chapter 17. There are also many diseases which impair specific immune responses; they include the rare congenital immunological deficiency states and also acquired diseases which involve the lymphoid tissues and affect their immunological functions (pp. 169–72).

In other diseases, predisposition to infection is well known but unexplained. A good example is diabetes mellitus, in which boils and urinary tract infections are common and there is a predisposition to tuberculosis: it may be that phagocytic activity, which requires the energy provided by glycolysis, is impaired by the defective carbohydrate metabolism of diabetes.

Two important reactions to infection are leucocytosis and fever, accounts of which follow.

Polymorphonuclear Leukocytosis

The number of neutrophil polymorphs in the blood, normally $2.5\text{--}7.5 \times 10^9/\text{litre}$ in older children and adults, increases in various pathological conditions. The increase is a controlled reaction and when the cause subsides the leucocyte count returns to the normal level for that individual. This account deals with the causes and mechanisms of such a neutrophil leukocytosis and with the changes in the haemopoietic marrow responsible for increased leucocyte production. The proliferation of leucocytes in myeloid leukaemia is neoplastic rather than reactive and is described in Chapter 16.

Causes

Neutrophil leukocytosis occurs in association with acute inflammatory reactions, tissue necrosis, thrombosis, haemorrhage, acute lysis of

red cells, and sometimes cancer. A mild polymorph leukocytosis occurs in pregnancy and also results from strenuous exercise, severe mental stress, and from injection of glucocorticoids, corticotrophin or adrenaline. By far the commonest cause in clinical medicine is inflammation due to bacterial infection, and in general the degree of leukocytosis correlates with the size of the inflammatory lesion and the intensity of polymorph emigration in the infected tissues. Pyogenic infections, due for example to virulent staphylococci, streptococci, pneumococci or coliform bacilli, are accompanied by a brisk leukocytosis, the height of which depends partly on the duration and partly on the extent of the infection. A boil or acute appendicitis may induce a moderate rise, e.g. to 10×10^9 polymorphs per litre, while a large abscess, acute bacterial pneumonia or general peritonitis are commonly accompanied

by a count of $20 \times 10^9/l$ or more. In some severe infections with pyogenic bacteria, e.g. streptococcal septicaemia or pneumococcal pneumonia, there may be absence of leukocytosis, or even leukopenia, and this usually indicates overwhelming toxæmia and is a bad prognostic sign. In some severe infections, e.g. gas gangrene due to *Cl. welchii*, polymorph emigration is less intense, and the increase in polymorphs in the blood is less marked, while the acute inflammatory lesions of the intestine caused by the typhoid and paratyphoid bacilli are virtually devoid of polymorphs (Fig. 7.4) and there is actually a fall in the number of polymorphs in the blood. Many virus infections, particularly in the early stages, are also accompanied by a neutrophil leukopenia.

Necrosis of tissue, for example myocardial infarction, causes a slight or moderate neutrophil leukocytosis, and extensive thrombosis, e.g. in the leg veins, or severe haemorrhage, both have a similar effect.

Leukocytosis may develop within a few hours of the onset of a bacterial infection and is of diagnostic value. This early rise is due partly to release of many polymorphs which normally lie margined in the venules of the lungs and elsewhere, and partly to release of polymorphs lying in the sinusoids of the haemopoietic marrow. Soon, however, there is an increased rate of formation of polymorphs in the marrow and the leukocytosis is thus maintained. It appears from recent observations that the life of the neutrophil polymorph in the

blood is probably not more than 12 hours, and since there are approximately 5 litres of blood containing about 4×10^9 polymorphs per litre, the normal daily production must be at least 4×10^{10} . In a suppurating infection, ten times this number may be lost daily for weeks or months in the pus discharging from an abscess, while at the same time the blood level may be maintained at $20 \times 10^9/l$ or more. It is thus apparent that the output of polymorphs is capable of enormous and sustained increase, and the process responsible for this is hyperplasia of the bone marrow, considered below.

Production of polymorphs (Granulopoiesis)

In the normal adult, production of the granulocyte or myeloid series of leukocytes is restricted to the haemopoietic marrow, where it occurs along with the production of red cells, platelets and monocytes. All these cells, and also lymphocytes (p. 118), originate from *haemopoietic stem cells*, which give rise to more stem cells and also to cells of more restricted potential: some are progenitors of red cells, others of megakaryocytes, while recent observations have demonstrated progenitor cells capable of giving rise to both polymorphs and monocytes (see below).

Because of their basic role in haemopoiesis, haemopoietic stem cells are dealt with in the chapter on blood (p. 504).

Stages of granulopoiesis. The earliest recognisable granulocyte precursor is termed a myeloblast: small numbers of these are present in normal haemopoietic marrow and they divide to give rise to a population of cells which undergo successive multiplications and form the largest cell population in the marrow. This proliferation is accompanied by a continuous process of differentiation up to the granulocyte stage: representative stages are illustrated in Fig. 7.5. Throughout the process the ratio of cytoplasm to nucleus increases and after initial enlargement up to the early myelocyte stage, diminution in size is a feature of differentiation. In the primitive stages the nucleus is large, ovoid or indented, and the chromatin is finely distributed. Gradually the nucleus shrinks, becoming more deeply staining, and eventually it becomes elongated giving the 'band form', followed by division into lobes, the number of which increases during the late stages of matur-



Fig. 7.4 Inflammatory infiltration of the muscular layer of the small intestine in typhoid fever, showing mononuclear cells and absence of polymorphs. $\times 150$.



Myeloblast: non-granular (oxidase-negative), basophilic cytoplasm (rich in RNA); nucleus roughly spherical, with dispersed chromatin (euchromatin), and containing 2 or more nucleoli. 10–18 μm diameter.

Promyelocyte: a few primary (oxidase positive) cytoplasmic granules; basophilic cytoplasm; nucleus spherical, still with dispersed chromatin and nucleoli. 12–18 μm diameter.

Myelocytes: cytoplasm less basophilic; primary granules disappear and secondary granules develop—they are strongly oxidase-positive and specific (neutrophil, eosinophil or basophil). Nucleus spherical or ovoid with some condensation of chromatin (heterochromatin). Active mitosis occurs at this stage. 12–18 μm diameter.

Metamyelocyte: cytoplasm only faintly basophilic (poor in RNA), with many specific granules and no primitive granules. Nucleus becoming elongated, smaller and more condensed. 12–15 μm diameter.

Polymorphonuclear leukocyte: cytoplasm as in metamyelocytes; nucleus smaller, and chromatin more condensed, at first horse-shoe shaped, later lobulated. 10–15 μm diameter.

ation of the polymorph in the marrow and in the blood. In preparations stained by a Romanowsky dye (e.g. Leishman's, Wright's, Jenner's or Giemsa stains), the cytoplasmic changes are as described in Fig. 7.5: from the myelocyte stage, there is not sufficient RNA to impart strong basophilia, and the cytoplasm is pale blue. Lysosomal granules appear in the cytoplasm in the promyelocyte stage: at first they are large and stained reddish-blue, but in the myelocyte stage these early granules are gradually replaced by granules specific for neutrophil, eosinophil or basophil polymorphs. In the neutrophil myelocytes the granules are small and are stained reddish or purple: in the eosinophil they are larger and bright orange, while in the basophil they are large and dark blue. These characteristic granules persist in the three types of mature granulocytes or polymorphs. In an early neutrophil leukocytosis there is an increased proportion of young cells and even myelocytes may appear in the blood.

Leukocytosis is brought about by hyperplasia, (i.e. an increase in the number of cells), in the haemopoietic marrow, and the proportion of myeloid cells, and particularly of myelocytes, is increased. This is termed a *granulopoietic reaction* and is analogous to the erythroblastic reaction in response to an increased requirement of red cells, e.g. after haemorrhage. The fat cells normally present in haemopoietic (red) marrow diminish in number as the cellularity increases (Fig. 7.6) and also foci of haemopoietic tissue arising from stem cells appear in the yellow fatty marrow of the long bones. These foci extend rapidly and in a severe prolonged suppurating infection much of the yellow marrow in the shafts of the femur and other long bones may be replaced by red marrow, the change starting in the upper ends of the bones and extending downwards. All the cellular constituents of normal marrow are present in this newly formed haemopoietic tissue, but myelocytes and later forms predominate (Figs. 7.7, 7.8).

Factors controlling granulopoiesis. It has long been known that many substances promote a neutrophil leukocytosis when injected into animals: they include peptones, digestion products of nucleic acids, and metabolic products and extracts of bacteria. Such observations have not helped much in the elucidation of the mechanisms of leukocytosis, but investigations

Fig. 7.5 Cells in haemopoietic marrow representing stages of granulopoiesis: from the myelocyte stage, only neutrophil cells are illustrated. Leishman stain $\times 1000$. (Dr. R. Brooke Hogg.)

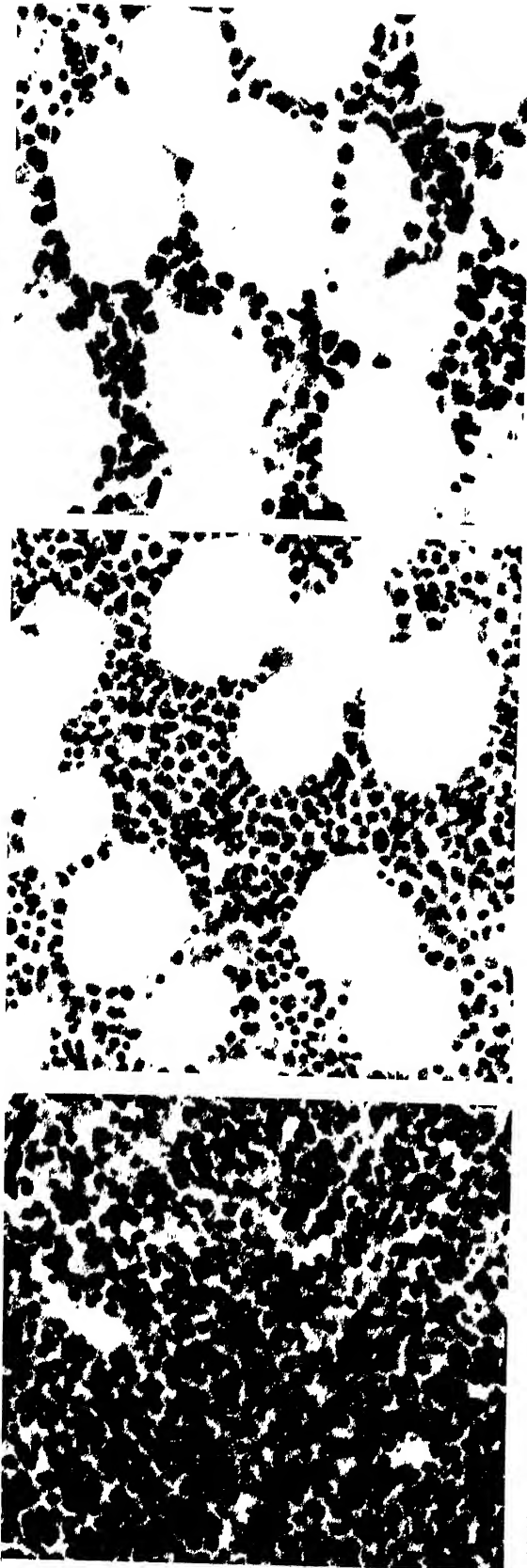


Fig. 7.6 Sections of haemopoietic marrow illustrating hyperplasia associated with neutrophil leukocytosis. *Top*, normal marrow. *Middle*, marrow showing increased cellularity in a patient with leukocytosis of short duration. *Bottom*, marked hyperplasia of marrow in a patient with prolonged leukocytosis: the fat cells have been replaced by haemopoietic cells. $\times 315$.



Fig. 7.7 Smear preparation of sternal marrow during a granulopoietic reaction. Note granular myelocytes in mitosis and various stages of transition to polymorphonuclear leukocytes. $\times 600$.

involving the culture of marrow cells *in vitro* have proved more successful. When cultured in a suitable semi-solid medium, the growth of myeloid cells requires the presence of a factor (colony stimulating activity or CSA) which stimulates the proliferation of individual cells to form discrete colonies. At first, the proliferating cells are mainly or entirely granulocyte precursors, but as the numbers increase some of the cells differentiate into polymorphs and others into monocytes. Since such colonies are derived from single cells, this is good evidence of a common precursor of polymorphs and monocytes. CSA can be extracted from most tissues (including haemopoietic marrow) and is present in the serum and urine. It is now known to be produced by monocytes, macro-

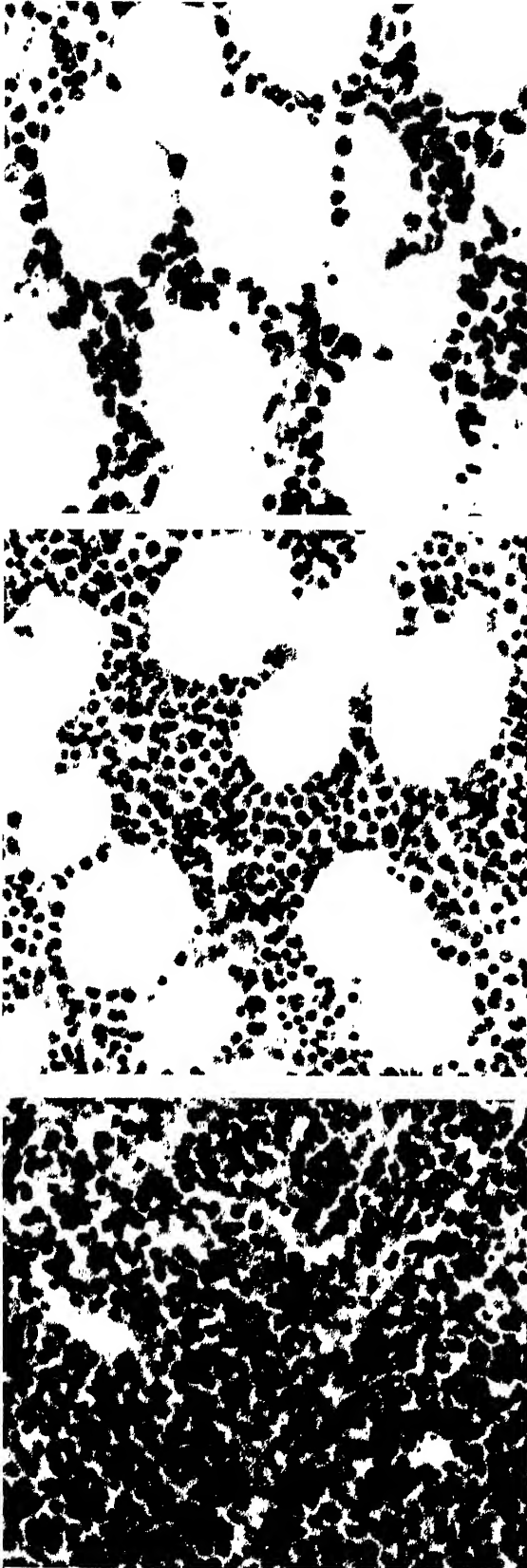


Fig. 7.7 Smear preparation of sternal marrow during a granulopoietic reaction. Note granular myelocytes in mitosis and various stages of transition to polymorphonuclear leukocytes. $\times 600$.

involving the culture of marrow cells *in vitro* have proved more successful. When cultured in a suitable semi-solid medium, the growth of myeloid cells requires the presence of a factor (colony stimulating activity or CSA) which stimulates the proliferation of individual cells to form discrete colonies. At first, the proliferating cells are mainly or entirely granulocyte precursors, but as the numbers increase some of the cells differentiate into polymorphs and others into monocytes. Since such colonies are derived from single cells, this is good evidence of a common precursor of polymorphs and monocytes. CSA can be extracted from most tissues (including haemopoietic marrow) and is present in the serum and urine. It is now known to be produced by monocytes, macro-

Fig. 7.6 Sections of haemopoietic marrow illustrating hyperplasia associated with neutrophil leukocytosis. *Top*, normal marrow. *Middle*, marrow showing increased cellularity in a patient with leukocytosis of short duration. *Bottom*, marked hyperplasia of marrow in a patient with prolonged leukocytosis: the fat cells have been replaced by haemopoietic cells. $\times 315$.



Fig. 7.8 Smear preparation of marrow showing finely granular myelocytes and transitions to polymorphonuclear leukocytes. $\times 1000$.

phages, vascular endothelium and stimulated lymphocytes. Assay of CSA by its effect on marrow culture is difficult and complex. This is probably because it is produced by monocytes in the culture, and in heavily seeded cultures sufficient CSA is produced to promote colony formation. It now seems likely that local production of CSA in haemopoietic marrow is important in the physiological production of

neutrophil polymorphs and monocytes. Mature neutrophils produce a factor (possibly lactoferrin) which inhibits leukopoiesis and thus provides a negative feedback mechanism. Shortly after an injection of bacterial endotoxin, a **polymorph releasing factor** appears in the plasma, and has the effect of releasing mature neutrophil polymorphs from the marrow sinusoids into the blood. It may be that, by thus reducing the production of the neutrophil inhibitory factor in the marrow, this allows hyperplasia with production and release of more granulocytes. Bacterial endotoxin also increases the production of CSA by monocytes in culture and *in vivo*, and so it may reduce the negative feedback in the marrow and also directly stimulate leukopoiesis.

Other factors of possible importance include a lipoprotein present in normal plasma which, when added to marrow cell cultures, is said to increase the production of monocytes at the expense of polymorphs. A factor is also produced by polymorphs in culture fluid which inhibits proliferation of early polymorph precursors in culture.

Although these observations promise considerable advances in our understanding of polymorph production, their relevance to natural neutrophil leukocytosis in man is still not known.

Pyrexia (Fever)

In this account, *pyrexia* and *fever* are used synonymously to mean a rise in the internal temperature of the body ('core temperature') to levels above the normal range. Traditionally, *fever* is also used in the nomenclature of various diseases (usually infections) in which pyrexia is a prominent feature, e.g. typhoid fever, yellow fever and cerebrospinal fever. Doubtless both these usages will continue.

Body temperature is controlled partly by reflexes initiated by the thermo-sensory nerve endings in the skin, and partly by a central control mechanism in the hypothalamus.

The peripheral reflex control mechanism can be demonstrated simply by observing a fall in the temperature of the skin of the left hand when the right hand is immersed in cold water.

This is brought about by impulses from the cold receptors in the chilled skin (of the right hand) which stimulate sympathetic vasoconstrictor fibres supplying the skin and subcutaneous tissues in general. The result is a reduction in heat loss and maintenance (or even rise) of the core temperature. This experiment works when the blood flow through the right arm is arrested by a pressure cuff, and so is not dependent on the temperature of the blood reaching the hypothalamus. In experimental animals, the reflex can be elicited even when the spinal cord has been transected at a higher level, and so is not attributable to sensory impulses reaching the brain.

The central thermo-regulatory mechanism may, for practical purposes, be likened to a

thermostat. The thermo-sensory centre, shown in animals to be in the anterior hypothalamus, responds to variations in the temperature of the blood flowing through it, and may be demonstrated experimentally by direct heating of the anterior hypothalamus, which results in a fall in the core temperature. Signals from the thermo-sensory centre influence the activity of other hypothalamic centres which regulate the physiological processes responsible for heat production and heat loss, thus controlling the core temperature. It is not known how the thermo-sensory centre responds to variations in local temperature. Experimental studies suggest that release of catecholamines and 5-hydroxytryptamine by nerve endings in the anterior hypothalamus are important in the control of temperature in extreme environmental conditions, and that there are marked species differences in the influence of these monoamines on body temperature. Under moderate environmental conditions, however, their depletion or inhibition does not seriously influence temperature control.

Fever accompanying infections and various other pathological conditions is attributable to a humoral effect on the hypothalamic thermo-sensory centre (analogous to the thermostat being set high). It is important to distinguish this from conditions in which the centre is functioning normally but, for various reasons, heat loss cannot keep pace with heat gain, and so the body temperature rises, e.g. during vigorous exercise in a hot, moist atmosphere.

Disturbances of the thermo-sensory centre

It has long been realised that injection of dead bacteria or bacterial products induces fever. A number of such products, termed **exogenous pyrogens**, have been detected in filtrates of cultures of various bacteria and fungi, but the *endotoxins* of Gram-negative bacteria have been most extensively investigated. On injection into rabbits, these phospholipid-polysaccharide-protein complexes or lipid A (p. 178) induce fever in about 1 hour, and this is followed by a refractory period in which further injections of endotoxin are ineffectual. Following injection of endotoxin or other exogenous pyrogen into rabbits, a second pyrogenic factor appears in the plasma; this causes fever in about 20 minutes following injection

into a second rabbit, and differs from endotoxin in being pyrogenic in rabbits rendered refractory to endotoxin. This second pyrogen, now called **endogenous** (or **leukocyte**) **pyrogen (EP)**, has also been detected in the plasma of animals in the early stages of febrile bacterial and viral infections. It is produced when suspensions of human or rabbit polymorphs or monocytes are stimulated by endotoxin or by readily phagocytosed material such as dead bacteria or antigen-antibody complexes, and when macrophages are activated by lymphocytes in delayed hypersensitivity reactions (p. 159). Synthesis of RNA and protein is necessary for EP production, which in polymorphs starts about 2 hours after stimulation: monocytes take longer but secrete much more EP than polymorphs, and the EPs produced by the two cell types differ in their molecular weights. Polymorphs and macrophages obtained from inflammatory exudates produce EP *spontaneously* when incubated in culture medium, indicating that they have been stimulated *in vivo*.

These various observations suggest that many of the agents capable of inducing fever act by stimulating production of EP. In addition to its production by human leukocytes *in vitro*, EP has been demonstrated in inflammatory exudates in man. Its detection in human plasma during infective fevers has proved difficult, but this is not surprising because fever can be induced in man by injection of amounts of EP too small to provide a detectable level in the recipient's plasma. EP has, however, been demonstrated in human plasma at the onset of a bout of malarial fever. Further investigations on EP would be greatly helped by a more sensitive method for its detection than that depending on production of fever in experimental animals.

The fever which accompanies tissue destruction, e.g. myocardial infarction, or necrotic tumours, is probably also mediated by EP released by phagocytes in the inflammatory reaction to the necrotic tissue, although it may be that tissue cells can also release pyrogens.

Effector mechanisms of fever (Fig. 7.9)

Although fever cannot be regarded as a physiological reaction, it is nevertheless brought about by stimulation of the physiological mech-

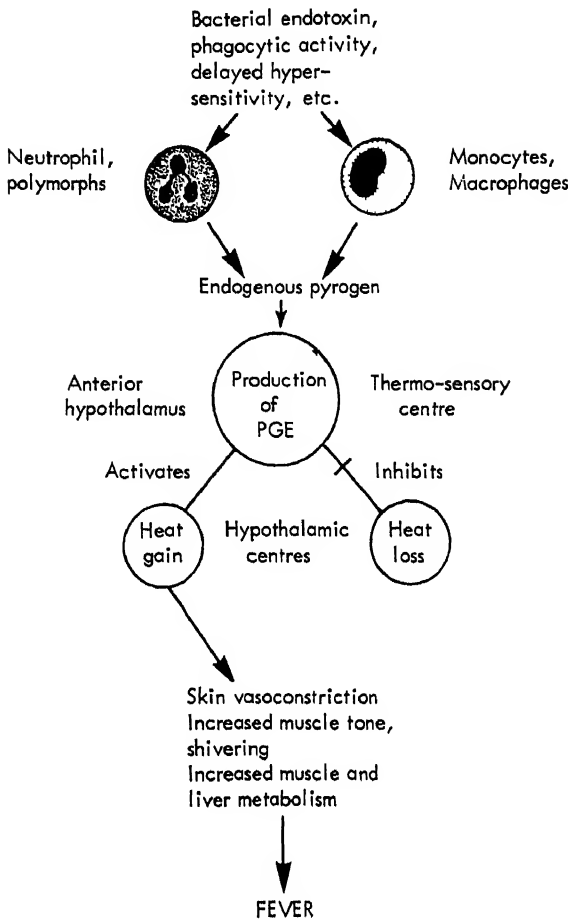


Fig. 7.9 The probable mechanism of fever in bacterial infections, etc.

animals for heat production and inhibition of those responsible for heat loss. For example, the shivering which accompanies a sharp rise of temperature is the normal response to a cold environment, and is associated with increased catabolic activity and heat production in the skeletal muscles. The coldness and pallor of the skin at the onset of fever are due to cutaneous vasoconstriction and, together with 'gooseflesh' (contraction of the pilo-erector muscles) and inhibition of sweating, are part of the normal heat-saving reaction to cold. Heat production is also increased in fever, as it is in a cold climate, by increased metabolic activity, particularly in the skeletal muscles (see above) and liver. This is mediated in part by stimulation of the sympathetic system and increased catecholamine secretion, and eventually thyroid activity may increase. A high caloric diet is

necessary to provide the fuel needed to maintain the body at temperatures above normal, failing which catabolism of endogenous fat and protein increases, resulting in a negative nitrogen balance and, especially in children, keto-acidosis. These catabolic activities account for the wasting commonly seen in patients with prolonged fever,* and the metabolic status in fever is, in fact, closely similar to that following injury (p. 265): in both, a warm environment and a high caloric intake, including increased protein, are beneficial.

Not only is the 'thermostat set high' in fever, but it is unstable, so that the temperature commonly fluctuates, and is readily affected by environmental conditions. As described above, a rise of temperature is achieved by increasing heat production and reducing heat loss by physiological mechanisms. Conversely, a fall of temperature, either during or at the end of a fever, is accomplished by reduction in catabolism and by cutaneous vasodilatation and sweating. The skin is flushed, warm and moist, and the patient feels hot. During fever, the degree of increased metabolic activity, etc. will depend largely on the environmental conditions and, in a chilling environment, on the degree of insulation of the body by clothing.

Heat production and loss are regulated by centres in the posterior hypothalamus, including one which influences activity of the sympathetic system, and a second which appears to control muscle tone and induction of shivering. When injected into the anterior hypothalamus, EP induces fever within minutes. It now appears likely that EP does not act on the thermo-sensory centre directly, but by stimulating the production of prostaglandins of the E series (PGE) in various parts of the brain, including the anterior hypothalamus. This view is supported by the following observations. (1) High levels of PGE are present in the cerebrospinal fluid during fever induced by endotoxin. (2) Injection of minute amounts of PGE₁ or E₂ induce fever in various species when injected directly into the anterior hypothalamus or third ventricle. (3) Aspirin-like drugs, which inhibit the synthesis of prostaglandins (p. 56) prevent the induction of fever by endotoxin (or by EP) but have no effect on fever induced by PGE. In the cat, PGE₂ is mainly involved, but other

* The old adage 'starve a fever, feed a cold' had little metabolic justification.

prostaglandins may be concerned in other species.

The role of prostaglandins explains the fever associated with induction of labour in pregnant women by an infusion of PGE₂; it would account also for the antipyretic effect of aspirin and similar drugs.

Cortisone is also antipyretic, but its inhibitory effect on production of EP by leukocytes stimulated by endotoxin, etc., probably accounts for this.

Effects of fever

The ill-effects of fever include general malaise, anorexia and increased catabolism. When the temperature rises to 41.6°C (107°F), there is a danger of direct thermal injury to various tissues, and particularly to cerebral neurones. In general, there is no evidence that fever has a beneficial effect, and its reduction by antipyretic drugs or by cooling the body does not seem to influence the course of infections. Apart from the spirochaete of syphilis and gonococcus (the cause of gonorrhoea), micro-organisms in culture do not appear to be adversely affected by moderate rises in temperature. The effects of fever on viruses are complex and require further investigation.

The spontaneous movement of neutrophil polymorphs *in vitro* and their response to a chemotactic stimulus are most rapid at 40°C (104°F), and fever may thus enhance the defensive role of these cells in infections.

Other causes of fever

Lesions of the hypothalamus may cause fever by interfering with the functioning of either the thermo-sensory centre or the hypothalamic areas which regulate heat loss and heat production. In experimental animals, injury of the anterior hypothalamus often causes pyrexia, while injury of the posterior hypothalamus may induce hypothermia. In man, haemorrhage in the pons is often accompanied by fever, and lesions between the hypothalamus and upper cervical cord interfere with tracts controlling heat loss and production, rendering the individual less able to respond to environmental temperature changes, etc.

Fever may occur in the absence of any disturbance of the thermo-sensory mechanism in

conditions where the physiological mechanisms of heat loss cannot keep pace with heat production. This occurs in *thyrotoxicosis*, in which excess secretion of thyroid hormone stimulates general metabolism and physical activity and thus increases heat production. In normal subjects, vigorous exercise or a hot moist environment may both cause fever, and the combination is particularly likely to do so. Obviously, heat loss is influenced by the temperature and humidity of the atmosphere, air currents and insulation by clothing. These factors affect loss of heat by conduction, convection, radiation and evaporation, and also by the cooling effect of inspired air. Sweating is a major mechanism of heat loss, but is only effective if the sweat evaporates on the skin surface, thus extracting the latent heat of vaporisation. Excessive sweating may, however, cause dehydration and, if water is restored, salt deficiency. Also, marked cutaneous vasodilatation may impair the circulation. These various factors are associated in combinations which give rise to several clinical syndromes, the chief of which are as follows.

1. **Heat exhaustion** results from physical activity in a hot climate, particularly if the atmosphere is moist. Vasodilatation in the skin and skeletal muscles creates a relative oligæmia, i.e. the filling of the enlarged vascular bed reduces the return of blood to the right side of the heart, and so cardiac output falls. Literally, there is not enough blood to go round. The heart rate increases and the blood pressure falls, giving a fast weak pulse, dyspnoea and other signs of circulatory insufficiency. The skin is hot and damp, and the subject feels tired and becomes confused. Rest and restoration of fluid usually bring rapid improvement.

2. **Dehydration exhaustion.** When dehydration due to excessive sweating, reduced fluid intake, etc., accompanies heat exhaustion, all the features of circulatory failure are exaggerated by actual, in addition to relative, reduction in the blood volume. Dehydration in a hot dry climate is also increased by loss of fluid through the epidermis, which is not entirely impervious. The core temperature may be very high and collapse and sudden death may occur.

3. **Heat stroke.** In the two conditions mentioned above, heat-losing mechanisms operate, but are inadequate. In heat stroke, exercise in a hot environment, with consequent fever, leads

in some way to a breakdown of the control mechanisms, so that the heat-losing mechanisms remain inactive, and the temperature continues to rise and may reach 43°C (109°F). At this temperature, brain injury is accompanied by coma and convulsions, and death or permanent brain injury results.

4. Heat cramps. Painful cramps in the muscles are the result of salt deficiency. This is liable to occur when there is excessive loss of water and salt by sweating and only the water is replaced.

5. Malignant hyperpyrexia. This is an unusual complication of general anaesthesia, usually with suxamethonium but also with other agents. During the anaesthesia, muscle tone increases and the temperature rises rapidly, often to above 42°C (107.5°F). Cyanosis, shock and keto-acidosis develop and death may occur from cardiac arrest unless the condition is recognised and treated. The underlying metabolic predisposition is not understood, but in some instances has been shown to run in families in which a raised level of plasma creatine phosphokinase has been reported.

Hypothermia

This may be defined as a fall in the core temperature of the body to below 35°C (95°F). It has no particular relevance to host–parasite relationships and is considered here simply because the processes involved in temperature control, described above in relation to fever, are equally important in hypothermia.

Hypothermia occurs when heat production fails to keep pace with heat loss. In robust adults, this occurs only in conditions of extreme heat loss, such as immersion in the sea or exposure on mountains. However, factors

which interfere with heat production predispose to hypothermia in less rigorous environmental conditions. Because of their relatively large surface area and thin layer of insulating fat, infants, particularly if premature, are especially liable to it. Old people, especially women, living in cold surroundings on an inadequate diet, are particularly prone to develop hypothermia in winter. Predisposing diseases include hypothyroidism, generalised skin diseases, psychiatric disturbances and conditions which impair consciousness, metabolic or physical activity, e.g. alcoholism, narcotic drugs, paralysis, severe trauma and general states of disability.

Pathology. In general, metabolic processes decline rapidly below 33°C (91°F) and the cardiac output, blood pressure and respiratory rate fall. Fluid leaks from the microvessels with consequent haemoconcentration and increased blood viscosity. Blood flow to the tissues is further impaired by peripheral vasoconstriction; hypoxia and CO₂ retention increase and a combination of respiratory and metabolic acidosis develops. Below 25°C (77°F), the thermoregulatory mechanism ceases to function, and death results from cardiac arrest.

The changes found at necropsy include venous thrombosis, multiple small infarcts in various organs, pulmonary haemorrhages and bronchopneumonia. Acute pancreatitis is a common complication in patients who survive.

Induction of hypothermia to reduce the metabolic requirements of the brain and other organs was formerly practised in surgical procedures involving interruption of the circulation, but it carries a risk of ventricular fibrillation and is now used only occasionally in association with a pump to maintain the circulation.

Further Reading

Mims, C. A. (1976). *The Pathogenesis of Infectious Disease*, pp. 246. Academic Press, London and New York; Grune and Stratton, New York.

See also bibliography for Chapter 8 (pp. 224–5).

Types of Infection

Within living memory, infective disease was the major cause of death throughout the world, and the elimination or reduction in the incidence of most of the important infections largely accounts for the greatly increased life-span in technologically advanced communities. Many factors have contributed to this decline of serious infections: they include improved standards of community and personal hygiene, better nutrition and housing, prophylactic immunisation and antimicrobial therapeutic agents.

In spite of these great triumphs, infective disease is still of considerable importance: it remains the major cause of death in many tropical and subtropical countries where, in addition to bacterial and viral infections, protozoal and metazoal parasites account for a great deal of illness. Even in countries where infections have been greatly reduced, many problems remain. The common cold is as common as ever, and upper respiratory virus infections are the major cause of absenteeism from school, office and factory. The rise in the volume and speed of

world travel has increased greatly the risk of epidemics of influenza, cholera, etc. Even antibiotics have not proved an unmixed blessing, for their use has resulted in the spread of resistant pathogenic bacteria, particularly in hospitals. There are, moreover, a number of important diseases which may eventually prove to be due to infections, for example rheumatoid arthritis, multiple sclerosis, ulcerative colitis and sarcoidosis. Virus infections may also prove to be important causal factors of the lymphoid neoplasms, including Burkitt's lymphoma and lymphatic leukaemia, and of other forms of cancer.

This chapter gives a brief account of the various types of infection and describes some of the more important examples. As with other forms of disease, the effects of infection depend not only on the nature of the lesion, but also on its site in the body, and for this reason the special features of infection of the lungs, kidneys, brain, etc., are described in the appropriate systematic chapters.

Virus Infections

Of all the pathogenic organisms which affect man, viruses show the most extreme degree of parasitism. In the extracellular state viruses are metabolically inert and depend absolutely on the metabolism of the host cell for their replication. Basically, all viruses consist of a protein shell or **capsid** surrounding and protecting the nucleic acid—which may be either DNA or RNA. The **virion** or complete infectious particle of some viruses has an additional outer layer, or **envelope**, partially derived from the plasma membrane of the host cell. When a virus enters a susceptible host cell, the nucleic acid is re-

leased from the capsid and becomes functionally active: it is either transcribed into messenger RNA, or acts itself as messenger RNA, which re-directs the synthetic pathways of the host cell to manufacture components for new virus particles. This involves the replication of nucleic acid molecules and the production of proteins which include both the non-structural proteins (e.g. enzymes) necessary for viral replicative processes and also the structural proteins which become incorporated in the capsid of new virus particles.

Apart from those viruses which enter the

host by the bite of an insect (e.g. yellow fever), or in the case of rabies virus by the bite of an animal, all parasitic viruses must enter the body by invading the surface epithelial cells of some part of the body. In many instances the site of initial infection is in the respiratory or alimentary tracts.

In man, most virus infections are mild and are followed by complete recovery. Many infections are entirely symptomless and immunity to reinfection is acquired without serious disturbance at the time of primary infection. Although latent infection with virus may continue for months or occasionally even for years, viruses do not form a non-invasive flora in the way that some bacteria do. A few virus infections, such as smallpox, regularly cause serious disease and even viruses such as *herpes simplex* or the enteroviruses, which generally cause mild or symptomless infection, may occasionally give rise to severe disease in an unusually susceptible host (Fig. 21.40, p. 756). Viral infections, especially of the respiratory tract, are extremely common in the community and are, in general, more frequent in childhood than in adult life.

Viruses are structurally simple parasites and do not produce disease by the elaboration of toxins as bacteria do. Lesions in viral infections are due to direct invasion of body tissues with subsequent cell damage due to the effect of viral replication in the host cells. In most clinically-apparent virus infections, replication of the virus is accompanied by death of the infected cell (Fig. 8.1). Some viruses induce fusion between infected and adjacent non-infected cells, with the formation of multinucleated giant cells, for example the Warthin-Finkeldey cell of measles (Fig. 18.8, p. 572). Such giant cells usually die, at least in tissue culture preparations, but their formation may be important in allowing virus to spread without entering the surrounding medium. There is increasing evidence that the immune response of the host may sometimes play an important role in causing lesions—for example in the development of bronchiolitis due to respiratory syncytial virus, which may be at least partly due to a hypersensitivity reaction in the lungs of the host. In arbovirus encephalitis, it has been postulated on the basis of some results of animal experiments that the lesions may be due to virus-antibody complexes inducing a type 3 hypersensitivity reaction rather than to the



Fig. 8.1 Part of a vesicular skin lesion in varicella, illustrating virus-induced cell injury. The virus has replicated in the epidermal cells, resulting in cell death. The resulting epidermal defect has then become distended with inflammatory exudate, forming the vesicle. Note the swelling and hyperchromatic nuclei of the colonised epidermal cells, lying free in the vesicle and in the underlying epidermis. $\times 96$. (The late Professor J. A. Milne.)

direct effect of the virus on the cells of the brain.

Unlike some bacterial infections, *virus diseases are not usually accompanied by a polymorphonuclear leukocytosis, but a lymphocytosis is common*. Most are associated with fever, and rash and lymphadenopathy are quite commonly seen. In the acute phase of virus infection a protein, **interferon**, can be detected in the blood and tissues. Interferon is released from cells in response to virus infection and when taken up by other cells makes them refractory to virus infection. Although the production of interferon is induced by virus, the protein itself is a species-specific cellular protein. It is not virus-specific in its antiviral effect but inhibits virtually all viruses. *Interferon production is an important host defence mechanism against virus infection and is probably the major factor in bringing about recovery from acute virus infections*. Although specific neutralising antibody is responsible for immunity to re-infection, it begins to appear in the bloodstream only when the acute infection is subsiding. It is notable that infants with immunological deficiencies resulting in impairment of T-lymphocyte function (p. 170 *et seq.*) are prone to develop chronic progressive vaccinia (vaccinia gangrenosa)

following vaccination, and chronic infection with measles virus has also been reported. This provides evidence that cell-mediated immunity plays an important role in limiting viral lesions in normal individuals.

Distribution of lesions in virus infections

In some instances, the main lesion is at the site of the initial infection. For example, the virus responsible for influenza gives rise to a localised infection which spreads rapidly throughout the epithelial lining of the larger air passages of the respiratory tract. This results in epithelial necrosis of varying extent, and the cell injury and loss results in acute inflammatory oedema, which is the major clinical feature of influenza. The severity of the illness depends on the extent of epithelial necrosis, but secondary bacterial infection of the damaged mucosa is also of importance, especially in major epidemics. Influenza virus may enter the bloodstream, but appears unable to replicate successfully in the cells of other tissues. Other examples of virus infections which remain localised, and produce lesions mostly at the site of initial infection, are molluscum contagiosum (p. 1054) and the common cold.

In many other instances, the initial infection is clinically silent, but the virus invades various other tissues and organs and produces characteristic lesions in them. Thus in varicella the initial infection is probably in the respiratory tract. From there the virus spreads widely, invading the blood (viraemia) and many other tissues and organs. The characteristic vesicular lesions in the skin are due to invasion of the epidermal cells, and are one manifestation of the systemic infection. Suppuration of the skin lesions (pustulation) is due to secondary bacterial infection. The childhood fevers, e.g. measles, mumps, rubella (and also smallpox) are other examples of generalised virus diseases which follow initial infection *via* the respiratory tract.

Because of the mode of virus spread in varicella, the incubation period between initial infection and appearance of symptoms is about 14 days, and it may be even longer in some other exanthemas, e.g. measles and mumps. *Poliovirus* also spreads in a complex fashion within the body: following ingestion of the virus, there is an initial infection of the Peyer's patches in the small intestine. The virus then

spreads to the regional lymph nodes, and in some instances produces viraemia. In a few individuals (e.g. about 1 per cent of those infected with poliovirus type 1) the organism invades the anterior horn cells of the spinal cord (Figs. 21.44 and 21.45, p. 759), causing paralytic poliomyelitis. The intestinal infection is clinically silent, but it nevertheless results in the development of antibody in the blood and in the appearance of IgA antibody in the gastrointestinal tract (p. 175): immunity is thus provided against subsequent infection with the same type of poliovirus.

Persistent virus infections are known to occur in man. Herpes simplex virus, for example, remains latent within the trigeminal ganglion but becomes activated from time to time, e.g. during pneumonia or other febrile illness, to produce vesicles around the mouth. Varicella virus also commonly remains latent, and may subsequently become active and replicate within the cells of the dorsal root ganglia to produce an attack of zoster (Fig. 21.43, p. 757).

There is considerable interest at present in **slow virus infections**, which may be defined as virus diseases having a long incubation period, in some instances years, and a prolonged and progressive course. Such diseases have been demonstrated to occur in certain animals, e.g. Aleutian disease of the mink, and it seems very likely that kuru (p. 761) and Creutzfeldt-Jakob disease are examples in man. The agent of scrapie, a widespread chronic disease of sheep, is most unusual in its remarkable resistance to heat and viricidal chemicals.

Active immunity can readily be produced by the administration of attenuated viruses, e.g. Sabin poliovirus vaccine, measles and yellow fever vaccines, and also—although somewhat less effectively—by inactivated viruses, e.g. influenza and rabies vaccines. Naturally-acquired immunity after virus infection is generally life-long and is due to the development in the blood of antibodies which neutralise the infectivity of viruses. However, in a few virus diseases, reinfections or repeated infections are common. This may be due to the existence of numerous serologically distinct strains of virus, e.g. the common cold, or to the virus undergoing antigenic variation, e.g. influenza. In the case of certain viruses, and especially herpes simplex and varicella-zoster viruses, reactivation of virus in the tissues despite the presence of cir-

culating antibody is not uncommon. The recurrences of infection are probably due to the ability of these viruses to remain latent within cells and to spread on re-activation directly through cell walls to infect neighbouring cells. The presence of antibodies in people who have

experienced a virus infection can be demonstrated by various *in-vitro* tests such as complement fixation, haemagglutination-inhibition and neutralisation tests.

The possible role of virus infections in neoplasia is considered in Chapter 11.

Acute Bacterial Infections

The several processes which constitute the acute inflammatory reaction have been described in Chapter 3. They are basically the same in all acute inflammatory reactions, including those due to bacterial infections, but they differ in detail depending on the properties of the causal agent and the special features of the tissue involved. In some lesions, for example, inflammatory oedema may be unusually severe, while in others emigration of polymorphs or fibrin deposition may be predominant. In consequence of these variations, some acute inflammatory lesions, usually due to infections, present sufficiently characteristic appearances to warrant the use of the following descriptive terms.

Catarrhal inflammation. Acute inflammation of a mucous membrane is accompanied by glandular secretion, usually of thin watery fluid. Injury to the surface epithelium, together with inflammatory exudation from the superficial underlying vessels, results in detachment of the epithelial cells, either singly or in sheets (Fig. 8.2), and the detached cells are carried away in the mixture of secretion and exudate. When the infection subsides, the epithelium is quickly restored by proliferation of surviving cells, although prolonged or recurrent catarrhal infections may result in formation of granulation tissue and eventually fibrosis, and alteration of the epithelium to a less specialised or sometimes to squamous type.

The best known example of acute catarrhal inflammation—the common cold or coryza—is initiated by virus infection of the nasal mucosa, but various pathogenic bacteria multiply on the inflamed mucosa and aggravate the inflammatory reaction. The mixture of secretion and exudate then becomes increasingly viscid and turbid due to emigration of increasing numbers of polymorphs until it may consist of



Fig. 8.2 Wall of bronchus in acute inflammation, showing desquamation of epithelium and polymorphonuclear leukocytic infiltration. $\times 250$.

a mixture of mucus and pus: the inflammation, initially catarrhal, thus becomes *muco-purulent*. Catarrhal bronchitis (Fig. 8.2) is seen in mild influenza and, as in the common cold, the initial virus infection is often complicated by bacterial infections, with similar effects. Bacterial infection also induces catarrhal inflammation, for example of the colon in bacillary dysentery of moderate severity, and chemically-induced inflammation, such as that induced by inhalation of formalin or other irritating gases or vapours, may also be catarrhal.

Pseudo-membranous inflammation. This is usually due to bacteria which have a low invasive capacity but grow on the surface of a

mucous membrane and produce exotoxins which cause superficial necrosis and acute inflammation of the underlying tissue. As the exudate passes to the surface, the fibrinogen in it clots within the necrotic surface layer. The fibrin and dead tissue together form the *false (pseudo-) membrane*, which contains also the causal bacteria, polymorphs and erythrocytes. Eventually the digestive activity of polymorph enzymes at the junction of living and dead tissue results in loosening and detachment of the pseudo-membrane; when the micro-organisms are destroyed, healing occurs, sometimes with some scarring. Examples of this type of inflammation are provided by diphtheria, usually affecting the pharynx or larynx (Fig. 8.3), and in the colon the more severe examples of bacillary dysentery.

Serous inflammation. This consists of acute inflammation in which there is copious fluid exudation but emigration of leukocytes and escape of red cells are minimal. The tissues become grossly oedematous and when the

lining membrane of a body cavity, e.g. the pleura, is involved, the 'serous' exudate accumulates in the cavity. Serous inflammation is seen in the early stages of many acute bacterial infections, and particularly in infection by *Clostridium oedematiens*, one of the causal organisms of gas gangrene.

The terms **fibrinous** and **haemorrhagic** are also applied to inflammation, to indicate respectively marked fibrin deposition and escape of red cells.

Two important variants of acute bacterial infections with special features—**pyogenic infections** and **gas gangrene**—are described below.

Pyogenic infections: suppuration

In many acute bacterial infections, emigration of polymorphs is intense, and these cells accumulate in huge numbers in the inflamed tissues. If, as commonly happens, tissue necrosis also occurs, then the dead tissue is digested and a cavity is formed which contains polymorph-rich (**purulent**) exudate, or **pus** (Figs. 8.4, 8.5). Such a cavity is called an **abscess** and the process of abscess formation is termed **suppuration**. The adjective **pyogenic** is applied to bacteria which cause suppuration. Pyogenic bacterial infection of a natural body cavity, such as a joint, the peritoneum or subarachnoid space, results in accumulation of pus in the cavity without the necessity for tissue necrosis and digestion.

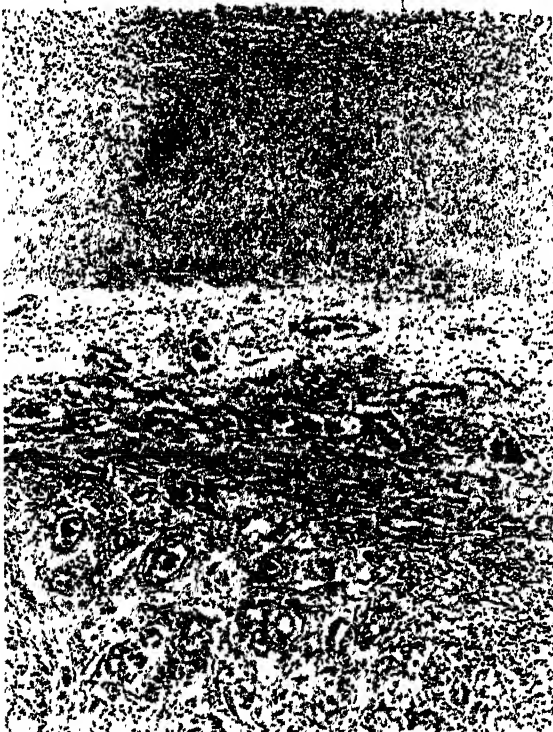


Fig. 8.3 Pharyngeal diphtheria. The mucosal surface (top) is coated with a false membrane composed of dead epithelium and fibrinous exudate. The underlying connective tissue shows acute inflammatory congestion. $\times 85$.

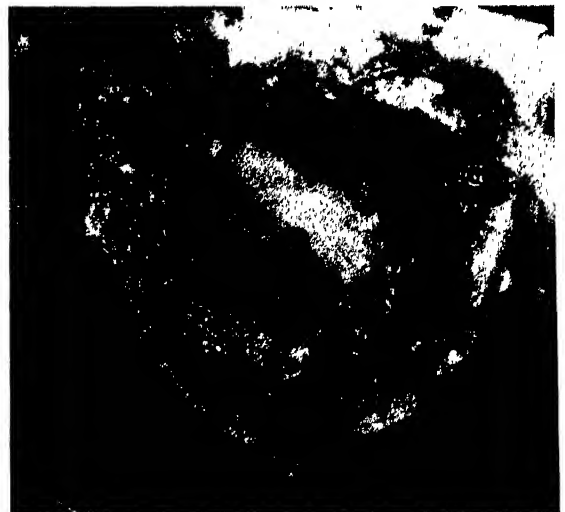


Fig. 8.4 Abscess of brain. Part of the skull has been removed surgically and a cavity containing pus is seen in the brain. (Photographed at necropsy.)

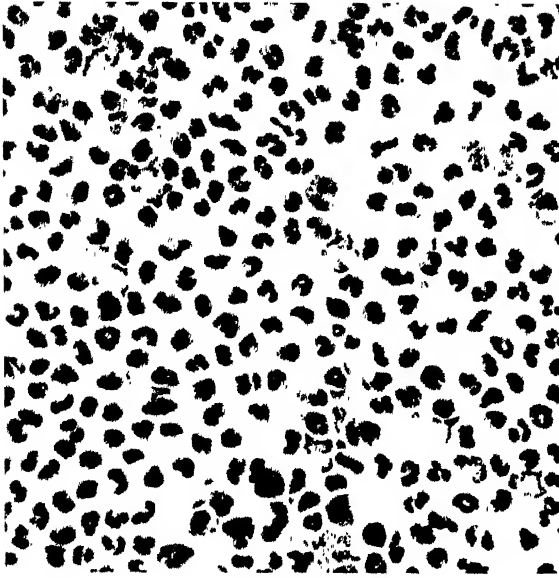


Fig. 8.5 Smear of pus. Most of the cells are neutrophil polymorphs: some are undergoing autolysis. $\times 400$.

Suppuration

Initially, a pyogenic bacterial infection shows the usual features of acute inflammation: as it progresses, local bacterial spread results in enlargement of the lesion, and unless the bacteria are destroyed rapidly, the tissue in the centre of the lesion undergoes necrosis. This is probably due mainly to the high concentrations of powerful toxins produced by pyogenic bacteria, but the pressure of inflammatory oedema, slowing of the blood flow, and sometimes thrombosis due to endothelial injury may also be important. The central necrotic tissue becomes infiltrated with polymorphs from the surrounding inflamed tissue, and during the process of phagocytosis and subsequent degeneration, these cells release lysosomal enzymes which digest the dead cells and tissue framework. Gradually a space, or abscess cavity, is formed containing fluid exudate rich in polymorphs, fragments of necrotic tissue, sometimes fibrin clots (from fibrinogen in the exudate), red cells and, of course, bacteria.

Necrosis of tissue and abscess formation favour multiplication of the causal bacteria. In acutely inflamed tissue, the continuous flow of exudate from small blood vessels into the tissue spaces and its removal by lymphatics is important in host defence (p. 62). In an abscess, however, the exudate is relatively stagnant: some

fluid can exude into the space from the surrounding inflamed tissue, but lymphatic drainage is inadequate. The intense migration of polymorphs into the abscess cavity also increases its contents, and in consequence the hydrostatic pressure in the cavity rises. The stagnant exudate in the abscess is a suitable growth medium for most pyogenic bacteria, and so they multiply and produce toxins which, by devitalisation of the surrounding living tissue, result in extension of the necrosis and enlargement of the abscess. Because of its raised pressure, the pus in an abscess tends to extend along tissue planes of least mechanical resistance; if present, for example, in the kidney, it may extend radially within and around the tubules, and may also burst through the capsule and spread extensively in the loose perinephric fatty tissue. An abscess forming near the skin, a mucous membrane or a serosal cavity, tends to extend towards the surface and rupture, discharging its pus.

If the growth of bacteria is checked, either by the natural defence mechanisms alone or with the help of antimicrobial drugs, the abscess stops enlarging, and becomes enclosed in a layer of granulation tissue (the *pyogenic membrane*) which grows from the surrounding inflamed tissue. Commonly, bacteria persist in the pus for a long time and the granulation tissue extends inwards while its outer part gradually matures to fibrous tissue. A long-standing abscess thus becomes enclosed in dense scar tissue which progressively thickens and, as long as the bacteria persist, is lined on its inner side by a layer of granulation tissue showing the changes of acute inflammation (Fig. 8.6). In the innermost granulation tissue, emigration of polymorphs may be conspicuous, while further out there may be plasma cells, lymphocytes and macrophages.

Extension of an abscess is accompanied by increase in toxæmia, fever and neutrophil leucocytosis, and rupture into a serosal cavity may result in extensive infection, e.g. generalised peritonitis or pleurisy. This is sometimes prevented, however, for as the abscess extends towards the cavity, fibrin in the inflammatory exudate glues adjacent viscera to the inflamed serosa, and by the time the abscess reaches the surface, that part of the cavity may be walled off. When an abscess ruptures through the skin or into the alimentary tract, the pus discharges



Fig. 8.6 Wall of an abscess. The abscess cavity is seen at the top right. The wall consists of vascular granulation tissue showing an inflammatory reaction. $\times 120$.

and the release of pressure allows free flow of exudate from the surrounding tissue into the abscess cavity: this favours elimination of the bacteria, not only by their removal in the discharging exudate, but also by the protective mechanisms afforded by a free flow of exudate (p. 62). Accordingly, surgical incision and drainage of an abscess is an important therapeutic measure: by promoting bacterial elimination and destruction, it allows the abscess cavity to heal with minimal scarring. The pressure in an abscess is well illustrated by the spurting out of pus when it is incised, but needs no emphasis for anyone who has experienced the throbbing pain of an apical tooth abscess.

Abscesses which are not drained, and which do not discharge naturally, may persist for months or even years and become surrounded by dense scar tissue. The bacteria may eventually be destroyed, and if the cavity is still small it may be gradually filled by granulation and eventually scar tissue. The pus in larger abscess cavities may be slowly transformed to clear fluid as the cell debris, etc. is removed by macrophages, leaving a cyst-like cavity which cannot collapse because of the surrounding rigid fibrous tissue. Occasionally the pus becomes inspissated to a solid, lipid-rich crumbly material, and deposition of calcium salts converts this into a stony hard mass which

may finally be replaced by bone. Even when an abscess is drained or discharges naturally, bacteria may persist in the cavity and drainage track, particularly if sufficient fibrosis has occurred to prevent its collapse.

Perhaps the commonest example of an abscess is a **boil (furuncle)**. It occurs most often in the dense dermal connective tissue at the back of the neck. The causal organism, *Staphylococcus aureus*, invades *via* the hair follicles or sebaceous ducts and sets up an acute inflammatory swelling. It spreads locally in the dermis, and necrosis of a patch of skin at the centre of the lesion results from toxic action and the vascular factors outlined above: polymorphs migrate from the surrounding inflamed tissue and digest the periphery of the necrotic 'core', which thus becomes separated from the lining tissue by a layer of pus. When separation is complete, the core is discharged, leaving an ulcer (Fig. 8.7). This is usually followed by elimination of the staphylococci, and the ulcer heals, leaving a pitted scar. In some instances, particularly in individuals with impaired resistance to infection, e.g. untreated diabetics, the infection may spread extensively in the dermal and underlying soft tissue of the neck, giving rise to a **carbuncle** consisting of a complex loculated abscess, or several separate abscesses, with multiple discharging sinuses (Fig. 8.8).

Suppuration in a serous cavity presents the same general features as an abscess developing in a solid tissue, and the principles of treatment are the same.

Composition of pus. As indicated above, pus consists of an accumulation of inflammatory exudate containing very large numbers of neutrophil polymorphs which give it an opaque appearance. Many of the polymorphs in recently formed pus are living, but the life of polymorphs which have emigrated is probably 12 hours or less, and in old pus most of the cells are dead and in various stages of degeneration and digestion. Release of DNA from these cells accounts for the sticky, slimy nature of pus. Some red cells are usually present, particularly in newly formed pus, and also fragments of tissue debris: fibrin may be present as free fragments, or may form a layer lining the wall of the cavity. In old pus, the number of macrophages increases and cholesterol crystals and globules of fat, derived



Fig. 8.7 Furuncle ('boil'). *Left*, the centre of the lesion is necrotic and about to be discharged. *Right*, the necrotic core has been discharged, leaving a ragged ulcer. $\times 1$. (The late Professor J. A. Milne.)

from degenerated cells and perhaps from blood lipids, gradually accumulate.

Bacterial infection of the blood

It is customary to classify the presence of bacteria in the blood into **bacteraemia**, **septicaemia** and **pyaemia**. The distinction between the three is not sharp, but they are none the less useful terms. In bacteraemia, bacteria are present in the blood in relatively small numbers but do not multiply significantly. Septicaemia and pyaemia are much more serious conditions in which bacteria, usually of high pathogenicity, multiply in the blood.

Only very rarely are bacteria present in the blood in sufficient numbers to be detected by direct microscopy, culture of the blood being necessary for their detection.

Bacteraemia. Small numbers of bacteria of low virulence are present from time to time in the blood of normal subjects, or in individuals with minor, often subclinical lesions. *Streptococcus viridans* may be cultured from the blood after vigorous brushing of the teeth, particularly if there is dental sepsis, and it is likely that occasional intestinal bacteria enter the portal circulation. Because of its high content



Fig. 8.8 Carbuncle: the foci of suppuration have extended to the overlying skin and discharged pus at several places. $\times 1$. (The late Professor J. A. Milne.)

of antibodies and complement, and the large numbers of circulating phagocytes and sinus-lining macrophages in the liver, spleen, etc., the blood is a hostile environment to most micro-

organisms, and although bacteria may multiply in local infections, those entering the blood are usually destroyed rapidly. Even in more serious and extensive localised infections, such as pneumococcal pneumonia or *Escherichia coli* infections of the urinary tract, bacteria can sometimes be detected by blood culture, but usually they fail to multiply significantly in the blood and disappear from it when, or even before, the local infection subsides. This applies also to the bacteria which enter the bloodstream as a regular feature of certain diseases, for example typhoid fever and brucellosis (undulant fever).

Bacteraemia is of some importance, for whenever they enter the blood, bacteria may settle in various parts of the body and cause lesions, for example suppurative meningitis or arthritis in pneumococcal pneumonia, and periostitis due to *Salmonella typhi* in typhoid fever.

Septicaemia means the presence and multiplication of bacteria in the blood, and is applied especially to the rapid multiplication of highly pathogenic bacteria, e.g. the pyogenic cocci or the plague bacillus, *Yersinia pestis*. The term thus implies a serious infection with profound toxæmia, in which the bacteria have overwhelmed the host defences.

In some instances it is difficult to distinguish between bacteraemia and septicaemia. For example, *Escherichia coli* causes infection of the peritoneal cavity, urinary and genital tracts: blood infection may occur, particularly as a complication of generalised peritonitis, but it is often not clear whether the bacteria are multiplying in the blood or are continuously entering it, e.g. from the infected peritoneum.

Multiple small haemorrhages may occur in septicaemia (Fig. 21.27, p. 747), due either to capillary endothelial damage from the severe toxæmia or to multiple minute metastatic foci of bacterial growth. The number of neutrophil polymorphs in the blood may be raised, although in overwhelmingly severe septicaemia they may be diminished and show toxic granulation (p. 512). The spleen is often enlarged and congested, and may contain large numbers of polymorphs. If the septicaemia is not rapidly fatal, foci of suppuration may develop in various parts of the body as a result of local invasion by blood-borne bacteria.

Pyæmia. In localised pyogenic infections, toxic injury to the endothelium of veins in-

volved in the lesion may result in thrombosis: bacteria multiply in the thrombus, which then becomes heavily infiltrated by polymorphs and broken down by their digestive enzymes. Small fragments of the softened septic thrombus may then break away and be carried off in the blood (**pyæmia**—literally, pus in the blood). Where they become impacted in small vessels, they cause local injury both by obstructing the vessels and by the release of toxins from their contained bacteria: a combination of necrosis, haemorrhage and suppuration results, with formation of multiple **pyæmic abscesses** in the various tissues, their distribution depending on the site of the original septic thrombosis. Pyæmic abscesses are typically surrounded by a zone of haemorrhage (Fig. 8.9). Microscopy of an early lesion may show a central zone of necrosis often containing huge numbers of bacteria (Fig. 8.10); this is surrounded by a zone of suppuration and an outermost zone of acutely inflamed, often haemorrhagic tissue. As the lesions progress, the necrotic tissue is digested, and apart from their multiplicity and widespread distribution, they become indistinguishable from non-haematogenous abscesses. In septic thrombosis of major veins, larger fragments may be released into the circulation, and



Fig. 8.9 The kidney in pyæmia, showing multiple small abscesses which are seen as pale areas surrounded by dark haemorrhagic zones.



Fig. 8.10 Pyaemic abscess of kidney in a case of staphylococcal pyaemia. Infarcted tissue (*above*) is separated from congested living tissue (*below*) by a zone of suppuration. The dark patches in a glomerulus are masses of staphylococci, and have probably increased after death. $\times 65$.

by impacting in arteries give rise to correspondingly larger foci of necrosis and suppuration (**septic infarcts**).

Septic thrombosis of systemic veins results especially, but not exclusively, in pyaemic abscesses in the lungs, while septic thrombosis in pulmonary veins results in pyaemic abscesses mainly in the systemic arterial distribution. In acute bacterial endocarditis, in which septic thrombus forms on the infected valve cusps, the distribution of pyaemic lesions depends on the particular heart valves involved. Septic thrombus of a portal venous tributary, e.g. in acute appendicitis, gives rise to **portal pyaemia**, with abscesses mainly in the liver.

Inevitably, bacteria are released from septic thrombus in pyaemia, and frank septicaemia commonly supervenes.

Septicaemia and pyaemia were formerly most often due to the pyogenic cocci. The incidence has, however, been greatly reduced by antibiotic therapy, and bacterial infections which are less readily eliminated by antibiotics have increased in relative importance, *Escherichia coli* now being the commonest cause of

blood infections in general hospital practice. In states of lowered resistance to bacterial infection, e.g. agranulocytosis, immunodeficiencies and therapeutic immunosuppression, blood infection is a particular hazard. Because of impaired defence mechanisms, bacteria normally of relatively low resistance may cause septicaemia and pyaemia in these conditions.

The common pyogenic bacteria

Pyogenic infections in man are most commonly caused by *Staphylococcus aureus* and *Streptococcus pyogenes*. *Staphylococcus aureus* is the usual cause of boils, carbuncles and septic lesions of the fingers: the infection usually remains localised, although suppurating lymphadenitis may occur in the local nodes, and unless treatment is effective, septicaemia and pyaemia may occur. Some types of *Staph. aureus* are resistant to penicillin and sometimes to other antibiotics. Symptomless nasopharyngeal carriers of resistant types are commonly responsible for outbreaks of infection of surgical wounds, burns, etc., in hospital: phage typing has proved of great value in tracing the source of such outbreaks. Staphylococci are also a cause of pneumonia complicating influenza and other virus infections of the respiratory tract, and may produce a fulminating enteritis in patients receiving broad-spectrum antibiotics. The staphylococcal lesion shows the usual features of acute inflammation, and unless checked by antibiotic therapy it frequently progresses to suppuration and discharges a thick creamy pus.

Streptococcus pyogenes commonly produces acute pharyngitis and tonsillitis, 'septic fingers', otitis media and mastoiditis, also extensive inflammation of the subcutaneous connective tissues (*cellulitis*), and *erysipelas*, a spreading infection of the dermis producing a raised, red, painful lesion of the skin, usually of the face, with a well-defined margin. Before the introduction of antiseptics, *Strep pyogenes* was a very common and important cause of fatal peritonitis or septicaemia arising from infection of the genital tract following childbirth. It can also cause fatal septicaemia resulting from a minor injury, e.g. a finger prick sustained by the surgeon or pathologist dealing with a streptococcal infection.

The differences between infections due to

staphylococci and streptococci are partly explicable by their toxins (p. 177). Staphylococcal infections show a greater tendency to remain localised, possibly due in part to the production of staphylocoagulase which clots fibrinogen, producing a deposit of fibrin which may help to limit spread of the organisms and promote their phagocytosis. Streptococcal lesions tend to spread, possibly due partly to the production of hyaluronidase, which digests hyaluronic acid and thus liquefies the ground substance of connective tissues. *Streptococcus pyogenes* also produces fibrinolysins, and leukocidins which kill polymorphs.

Other pyogenic bacteria include *Strep. pneumoniae* (the common cause of lobar pneumonia) which may be complicated by metastatic blood-borne lesions, e.g. suppurative meningitis or arthritis; *Neisseria meningitidis* (meningococcus) which invades the nasopharynx, often silently, and produces a septicaemia or bacteraemia with the subsequent development of meningitis; *Neisseria gonorrhoeae* (gonococcus), transmitted by coitus and producing an acute urethritis, etc.

The intestinal commensals are important causes of pyogenic infections in the abdomen, e.g. appendicitis, diverticulitis and peritonitis, and in the lungs; they also infect surgical and other wounds, bedsores, burns and ulcers of the skin. They include *Escherichia coli*, *Bacteroides*, anaerobic streptococci, *Proteus*, *Pseudomonas pyocyanea* and *Klebsiella*, and may cause infections singly and in various combinations. These organisms are of particular importance in debilitated and immunosuppressed patients. All of them, but especially *Esch. coli* and *Bacterioides*, give rise to septicaemia and pyaemia, with severe septic shock (p. 264).

Gangrene

Definition. The term gangrene means digestion of dead tissue by saprophytic bacteria, i.e. bacteria which are incapable of invading and multiplying in living tissues. Many types of bacteria, often present in various combinations, may participate, and breakdown of tissue proteins, carbohydrates and fat may result in simple end-products: volatile products and gases may be formed, giving the foul odour of putrefaction, and the same changes are observed in putrefaction of meat, etc. Gas pro-

duction may give rise to emphysematous crackling on palpation. The changes in colour—dark-brown or greenish-brown, and sometimes almost black—are due to changes in haemoglobin, and are most conspicuous when the dead tissue contains a lot of blood.

Gangrene may be either *primary* or *secondary*. The difference lies in the cause of the tissue necrosis. In primary gangrene it is brought about by the toxins of bacteria (which may then invade and digest the dead tissue). In secondary gangrene, necrosis is due to some other cause—usually loss of blood supply from vascular obstruction or tissue laceration—and saprophytic bacteria then digest the dead tissue.

Primary gangrene

This includes *gas gangrene* which results from infection with specific *clostridia*, and gangrene brought about by various other bacteria.

Gas gangrene is caused by a group of anaerobic sporulating bacteria, the *Clostridia*, of which the three most important are *Cl. welchii*, *Cl. oedematiens* and *Cl. septicum*. These organisms are intestinal commensals in man and animals; their spores are widespread, and are liable to contaminate wounds. Being anaerobic and saprophytic, they cannot multiply in living, oxygenated tissue, but they flourish in blood-soaked foreign material and dead tissue in dirty puncture or lacerated wounds such as are caused by road accidents and by shrapnel. Given such a favourable environment, the *Clostridia* produce exotoxins which diffuse into and kill the adjacent tissues and these in turn are invaded, so that the process spreads rapidly, particularly along the length of skeletal muscles (Fig. 8.11). Gas gangrene is most often due to *Cl. welchii*. Before the muscle and other tissues are killed, they become intensely oedematous, are extremely painful, and appear swollen and pink. Microscopically, emigration of leukocytes is minimal. Among a number of toxins, *Cl. welchii* produces a lecithinase (α toxin) which by its action on phospholipids lyses cell and mitochondrial membranes, also hyaluronidase and collagenase which digest ground substance and collagen respectively and may promote the rapid spread of infection. *Cl. welchii* ferments sugars, producing H_2 and CO_2 which collect as bubbles in the dead tissues,

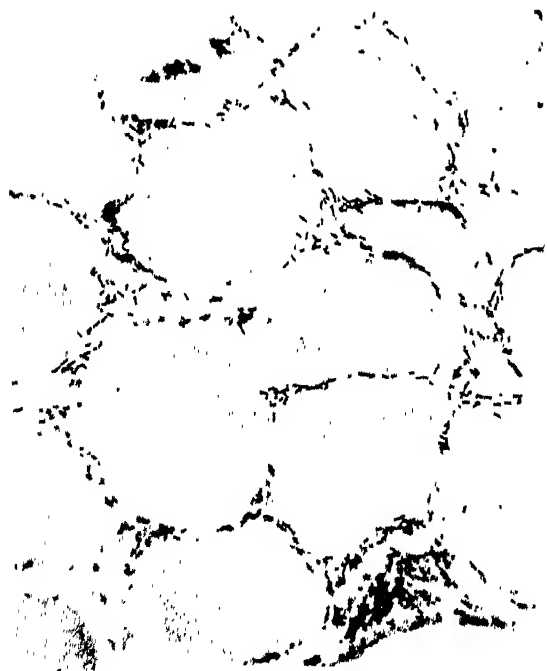


Fig. 8.11 Skeletal muscle in gas gangrene. The muscle fibres are necrotic and their nuclei have disappeared. Large numbers of *Cl. welchii* are present in the dead tissue, mainly in the endomysium. $\times 400$.

rendering them crepitant on palpation. The subcutaneous tissue and skin are also involved, and the affected part, often a limb, may burst open as a result of the swelling of oedema fluid and pressure of gas. The dead tissues are commonly invaded by a mixture of other organisms present in the original wound, and these may play a major role in putrefaction. Gas gangrene may complicate intestinal lesions, e.g. appendicitis or strangulation of the gut (p. 643), clostridia being present in the intestine; it occurs also as a puerperal infection of the uterus from contamination via the perineum.

In addition to the rapidly spreading local lesion, gas gangrene is accompanied by acute haemolysis and a severe toxæmia which affects all the internal organs, and death results from peripheral vascular collapse. Spread by the blood stream may occur as a late, usually terminal event. In consequence all the tissues at necropsy may contain large numbers of clostridia, with extensive digestive changes.*

The clostridia of gas gangrene may also

cause infection of subcutaneous tissue (cellulitis) without affecting the underlying muscle, or may grow in dirty wounds, producing a foul discharge but without either severe toxæmia or invasion of the surrounding tissues.

Production of toxin by *Cl. welchii* growing in contaminated food is also one cause of toxic gastroenteritis which is self-limiting and does not involve clostridial invasion of the gut wall.

If lacerated wounds are treated by early excision of the devitalised tissue, and toxic action is controlled by the use of antitoxic sera, clostridial infections are unlikely to establish themselves. Antibiotics have also contributed greatly to the prevention of gas gangrene by inhibiting the growth of clostridia.

Other examples of primary gangrene. In gas gangrene we have an example of primary gangrene caused by members of a group of clostridia whose toxins kill the tissue and, sometimes in association with other anaerobes, digest it. Primary gangrene can be caused by many other bacteria and mixtures of bacteria. For example, inhalation of dirty water in partial drowning, or of foul discharge from an ulcerated and infected cancer of the larynx, or from an oesophageal cancer which has ulcerated into the trachea, can each cause a gangrenous infection of the lungs in which various bacteria and fungi participate.

Gangrenous infection is particularly liable to develop in debilitated individuals and in those whose resistance to infection is lowered by various diseases; for example gangrenous pharyngitis or colitis may develop in patients with agranulocytosis, who lack the defences provided by neutrophil polymorphs. Diabetics also, unless their carbohydrate metabolism is adequately controlled, are particularly susceptible to infections, and gangrene may supervene. The following two unusual forms of primary gangrene also deserve mention.

'Meleney's post-operative synergistic gangrene.'

This is a slowly spreading infection of the skin and subcutaneous tissue of the chest or abdominal wall; it starts at the site of an operation wound and usually the operation has been performed to deal with a focus of sepsis in the chest or abdomen. The spreading edge of the lesion is acutely inflamed and

* When necropsy is performed after a body has remained for a day or more in a warm environment, putrefactive changes, including bubbles of gas, may be seen in various tissues and especially in the liver and other abdominal viscera. This is due to agonal and post-mortem spread and growth of *Cl. welchii* etc. present in the gut. It is seen especially in obese or oedematous subjects dying from any cause and must not be confused with gas gangrene.

appears red and swollen: as it spreads, the central zone becomes darker and finally gangrenous, followed by sloughing to leave an ulcerated area with a granulating base. The lesion may spread relentlessly to involve much of the trunk. It is usually caused by a synergistic combination of *Staphylococcus aureus* and a streptococcus, but other combinations may have a similar effect.

Noma (cancrum oris) is a gangrenous condition occasionally seen in poorly nourished children and tends to complicate debilitating infections; it begins on the gum margin and spreads to the cheek, where an inflammatory patch of dusky red appearance forms and then becomes darker in colour and ultimately gangrenous. The condition is caused by bacteria of the genus *Bacteroides* (anaerobic bacilli present in huge numbers in the intestine and also part of the normal flora of the mouth) together with *Borrelia vincenti*, another mouth commensal. Deficient intake of the vitamin B complex, especially of nicotinic acid, is a predisposing factor.

Secondary gangrene

This is usually the result of ischaemic necrosis (from loss of blood supply) followed by invasion and digestion of the dead tissue by putrefactive micro-organisms. It is seen most often in the foot and leg, and in the intestine. As explained below, it occurs in two forms—'wet' and 'dry' gangrene.

Gangrene of the leg. Infarction of toes, a foot, or the lower leg is not uncommon as the result of arterial occlusion (Fig. 2.6, p. 12), the collateral circulation being insufficient to keep the part alive. This is caused by arterial thrombosis complicating advanced atheroma (p. 238), which is very common in old people and tends to be particularly severe in diabetics (hence the terms '*senile*' and '*diabetic*' gangrene). It may occur also in early or middle adult life in patients with thrombo-angiitis obliterans, a disease which affects multiple arterial branches, especially in the lower limbs. Another occasional cause is the symmetrical spasmodic contraction of arteries in Raynaud's disease.

If there is much subcutaneous fat, and particularly when the limb is oedematous, as in congestive heart failure, **wet gangrene** commonly supervenes in the infarcted tissues, with blebs of fluid in the skin, sometimes gas production, and rapid putrefaction: there is no sharp line of demarcation between dead and

living tissue, and indeed gangrene may spread proximally beyond the tissues originally affected. When infarction occurs in a non-oedematous leg, particularly when there is little subcutaneous fat and when gradual arterial occlusion has preceded the actual infarction, so-called **dry gangrene** is liable to ensue: the skin becomes cold and waxen, the haemoglobin diffuses out of the veins and produces reddish-purple staining of the dead tissues, which then become brownish-red and ultimately almost black, and the dead tissue gradually dries out and shrinks (*mummification*).

Use of the term dry gangrene is controversial. Commonly, mummification occurs with little or no putrefaction. Saprophytic organisms are, however, usually present in small numbers, particularly adjacent to the junction with living tissue, where the dead tissue remains moist. If amputation is not performed, putrefaction becomes established at this site, and a process of slow putrefactive ulceration penetrates the soft tissues, ultimately down to the bone.

Secondary gangrene of the intestine occurs when the blood supply to part of the intestine is arrested by thrombosis of the mesenteric arteries or when a loop of intestine becomes impacted in a hernial sac. In the latter case, secretion of fluid and gas production by bacteria in the lumen result in a rise of pressure in the entrapped loop, with consequent interference with blood flow. In both cases, the impaired blood flow results in necrosis of the wall of the intestine, which is invaded by putrefactive bacteria from the lumen, becomes gangrenous, and ruptures unless removed without delay. While the intestine is dying, the wall becomes swollen and at first red and then black from fluid and red cells escaping from the small vessels. The features are thus those of wet gangrene. Ischaemic necrosis of other organs, e.g. the pancreas, may also progress to wet gangrene if it becomes infected with putrefying bacteria from the gut.

Anthrax

Anthrax is a fatal epizootic disease of animals, particularly cattle and sheep, caused by a large Gram-positive sporulating bacillus, the spores of which can survive for 50 years or more in soil.* In herbivores the disease is contracted by

* The island of Gruinard, off the West Coast of Scotland, is still contaminated with anthrax spores deposited experimentally during the 1939-45 war.

ingestion of spores and causes a severe acute enteritis with a terminal septicaemia: the excreta and secretions are highly infective. More chronic, localised lesions can also occur in animals. In man, *B. anthracis* is of low infectivity, but acute lesions occur in the skin from direct contact with infected material, or more rarely internally from inhalation or ingestion of spores.

The factors determining the virulence of the bacillus include a capsular polypeptide rich in D-glutamic acid, which renders the organism resistant to phagocytosis, and a complex exotoxin which promotes increased vascular permeability, causing gross inflammatory oedema. Death can result from hypovolaemic shock due to local and generalised exudative loss of plasma fluid.

Cutaneous anthrax (malignant pustule) of man occurs from direct contact with animal material, e.g. carcasses, hides or bristles in shaving brushes. Although many imported hides are contaminated with spores, anthrax is rare among those handling them. The organism probably enters through a minor abrasion, and a painful papule forms and becomes blistered:

it is surrounded by a zone of intense congestion and oedema. Central haemorrhage and necrosis follow, resulting in a black crust (Fig. 8.12). Leukocytic emigration is usually scanty. Spread may occur to the regional lymph nodes, which become enlarged, oedematous and haemorrhagic. Although uncommon, the condition is an important example of a serious, sometimes fatal infection which can be effectively treated if diagnosed early.

Respiratory anthrax occurs from inhaling spores, usually from hides or wool. A localised lesion develops in the lower trachea or larger bronchi: it consists of a patch of haemorrhagic, ulcerated mucosa with intense oedema, involvement of hilar and mediastinal lymph nodes, extension to the lungs and haemorrhagic pleural and pericardial effusions: the prognosis is poor.

Intestinal anthrax is rare in man. It consists of one or more haemorrhagic foci in the wall of the upper small intestine, with central necrosis, gross oedematous swelling and involvement of the mesenteric lymph nodes.

Septicaemia and a haemorrhagic meningitis may occur in man, but are rare.



Fig. 8.12 Anthrax. *Left*, before treatment. The malignant pustule is seen on the forehead. Note the gross and extensive inflammatory oedema. *Right*, following treatment. (By kind permission of Dr. W. M. Jamieson and the Editor of *Medicine*.)

Chronic Bacterial Infections (Infective Granulomas)*

Under this heading may be included the many infections which give rise to chronic inflammation without a conspicuous exudative reaction, but usually with production of granulation tissue which eventually progresses to fibrosis. The general features of chronic inflammation have been described on pp. 67–71 and a more detailed account of the production of granulation tissue is given on pp. 80–7. The following account describes briefly some of the more important chronic infections. Because of its wide prevalence and tuberculous-like features, sarcoidosis is also included here, although there is no evidence that it is an infection.

Tuberculosis

Formerly one of the great killing diseases of temperate climates, tuberculosis is now much less common in Western Europe and North America. It is, however, prevalent in communities with a poor standard of living, and still ranks among the world's most important diseases. The disease illustrates well various basic features of bacterial infection, and in particular the importance of the reaction of the host in determining the nature of the lesions, and the spread of infection within the body. The causal mycobacteria, or tubercle bacilli, are aerobic Gram-positive bacilli with a waxy cell wall which renders them difficult to stain. Once the stain has penetrated the cell wall, however, it is also difficult to remove, and the mycobacteria are sometimes referred to as acid- and alcohol-fast bacteria, since they resist decolourisation by mineral acids and alcohol, as in the Zeihl–Neelsen stain. Tubercle bacilli grow slowly in culture; they are highly pathogenic for the guinea-pig, inoculation of which has been much used for their detection when present in small numbers in sputum, etc. In man, they usually cause chronic disease but can also produce a much more acute and even rapidly fatal infection. In addition to the tubercle bacilli, of which there are two major types causing human disease (see below), the

mycobacteria include the lepra bacillus which causes leprosy, and there are also various ill-defined organisms, sometimes termed *anonymous* or *atypical mycobacteria*, which cause lesions in the skin, lymph nodes, lungs and elsewhere.

Epidemiology

The two types of tubercle bacillus mainly responsible for disease in man are the human type, *Mycobacterium tuberculosis*, and the bovine type *Mycobacterium bovis*. The *human type* is the more important: infection with it is usually contracted by inhalation, and the initial or primary lesion is nearly always in the lungs. Patients with chronic pulmonary tuberculosis provide the reservoir of infection and spread the disease by exhaling infected droplets and by coughing up infected sputum. The organism is resistant to drying and can survive for long periods in dust, inhalation of which is the usual method of contracting the disease. Infection of the tonsils or of the intestine can also occur from swallowing the human type of tubercle bacillus in contaminated dust, or the bovine type of bacillus in contaminated milk from cows with tuberculous mastitis.

Several factors are responsible for the declining incidence of tuberculosis in Western Europe and North America. Firstly, the rising standard of nutrition and housing: there is no doubt that under-nourishment predisposes to tuberculosis and impairs the resistance of the individual who has contracted the disease. Overcrowding and inadequate personal and domestic hygiene are also of importance in spreading the disease in the home and in public transport and meeting places, etc. The environment of a subject coughing up the organism is likely to be heavily contaminated, and spread within families is especially common, giving rise to both pulmonary and alimentary infections.

Since the 1939–45 war, the use of specific chemotherapeutic bactericidal agents has also helped to reduce the incidence of the disease

* As stated on p. 70, there is an increasing tendency to restrict *granuloma* to inflammatory lesions consisting of aggregates of macrophages. I have preferred to use *macrophage granuloma* to describe such a lesion, and to retain the traditional usage of *granuloma* to mean any chronic inflammatory lesion.

by diminishing greatly the infectivity of patients with chronic pulmonary tuberculosis. Mass miniature radiography has revealed unsuspected cases of tuberculosis in the community, and protection against infection has been provided by means of BCG vaccination.

In countries where the disease is rife, infants and young children are particularly at risk, and in this country the mortality rate in children contracting the infection before the age of 3 years was formerly very high. Those who overcome the infection develop partial resistance to the organism, but may become re-infected and develop chronic pulmonary tuberculosis in adult life. The bacteria may survive for many years in dormant lesions, without clinical manifestations, and these may become active as a result of malnutrition, as in war or famine, as a complication of other debilitating diseases such as diabetes mellitus, or from administration of corticosteroids or other immunosuppressive agents. In Western Europe, a high proportion of 'new' cases are middle-aged or old and have had dormant lesions for many years from the time when the disease was much commoner. It is, however, relatively common among Asian immigrants, younger age groups being affected. *Mycobacterium bovis* causes mastitis in cattle, and is transmitted to man by consuming infected milk and milk products. Infection results usually by way of the gut or tonsils. In many countries, bovine infection in man has been eradicated by pasteurisation of milk, which kills the organism, and by tuberculin testing of cattle and elimination of infected cows. The avian tubercle bacillus (*Myco. avium*) is a rare cause of disease in man.

Hypersensitivity and immunity

The immune response to the tubercle bacillus provides the classical example of cell-mediated immunity, i.e. the production of specifically primed T lymphocytes which are capable of reacting directly with antigenic protein of the mycobacterium. The mechanism of this type of response, and the state of delayed hypersensitivity which results from it, have been described in Chapters 5 and 6 respectively. It is not understood why cell-mediated immunity is the dominant type of immune response to the

tubercle bacillus, but it may be of significance that mycobacteria, living or dead, have a powerful enhancing effect on the cell-mediated immune response to antigens in general, and this forms the basis of their use in Freund's adjuvant (p. 114). Whatever the explanation of its adjuvant effect, infection with tubercle bacilli results, within two weeks or so, in the development of a high degree of cell-mediated immunity to a protein fraction (tuberculo-protein) of the organism* and the subsequent course of the infection and the features of the lesions are profoundly influenced by the hypersensitivity state. The specifically primed T cells react with tuberculo-protein, and release the various lymphokines described on p. 158. The results are both beneficial and harmful. *The tubercle bacillus has not been shown to produce any direct toxic effect, and can survive and multiply within macrophages in tissue culture without harm to the cultured cells. Indeed, it is likely that the tissue injury resulting from tuberculous infection is due mainly or entirely to the delayed hypersensitivity reaction against the bacteria.* Nevertheless, without an immune response, multiplication of the organism would presumably continue unchecked. The delayed hypersensitivity reaction is therefore to be regarded as protective in reducing or eliminating the infection, but at the same time injurious to the tissues. Interpretation of the features of the lesions of tuberculosis in terms of delayed hypersensitivity is attempted in the account of structural changes (below).

Although antibodies to mycobacterial antigens develop in tuberculosis, they do not appear to influence the course of the infection, and have not provided a useful diagnostic test.

Tuberculin skin testing. This is carried out by intradermal injection of very small amounts of tuberculo-protein, as in the *Mantoux test*. In individuals who are, or have previously been, infected, a delayed hypersensitivity reaction develops, the features of which are described on p. 157. In some patients with very severe tuberculosis, the test is negative, presumably because the large amount of tuberculo-protein being released from the lesions has overwhelmed the state of hypersensitivity. Tuberculin skin tests give positive reactions in infections with both

* This state of hypersensitivity was demonstrated by Robert Koch (1891), using a crude preparation termed 'old tuberculin'. A more refined preparation is termed 'purified protein derivative' (PPD).

human and bovine types of tubercle bacillus, and with other types of mycobacteria. For this reason they may be positive in individuals infected with *Mycobacterium leprae*. Positive tests are also observed in healthy inhabitants of tropical and sub-tropical countries, apparently as a result of previous sub-clinical infection with other mycobacteria.

Immunisation against tuberculosis. Protective immunisation requires the induction of cell-mediated immunity to tuberculo-protein, and this is most effectively achieved by injecting living mycobacteria. Attenuated strains of the bovine type, e.g. *bacille Calmette-Guérin* (BCG), or other non-human strains such as *Myco. muris* (the vole bacillus), are used for this purpose. Cell-mediated immunity develops, with consequent delayed hypersensitivity reactions at the site of injection and sometimes also in the draining lymph nodes which may become infected. The attenuated bacilli are destroyed and the lesions heal, but the cell-mediated immunity persists.

Structural changes

When a guinea-pig is inoculated with *Myco. tuberculosis* there is little reaction during the first day or so apart from local infiltration with neutrophil polymorphs, which soon disappear. During the next few days, macrophages migrate into the area and ingest the bacteria without bringing about their destruction. After ten days or so, lymphocytes begin to appear in the lesion, and macrophages derived mainly from monocytes of the blood aggregate in increasing numbers to form a minute nodule consisting of a macrophage granuloma. These very early stages of infection cannot, of course, be observed in man, but the subsequent changes are closely similar in man and the guinea-pig. The macrophages enlarge and change to **epithelioid cells** (p. 74). Small lymphocytes accumulate around the margin of the nodule, which is then termed a **tubercle** and becomes visible to the naked eye about 3 weeks after the onset. In the central part of the lesion, multinucleated **Langhans' giant cells** (p. 74) are formed by fusion of epithelioid cells (Fig. 8.13). As the tubercle enlarges, the epithelioid and giant cells in the central part undergo necrosis (Fig. 8.14): the cells lose their outline and nuclear staining and become fused into a homogeneous or

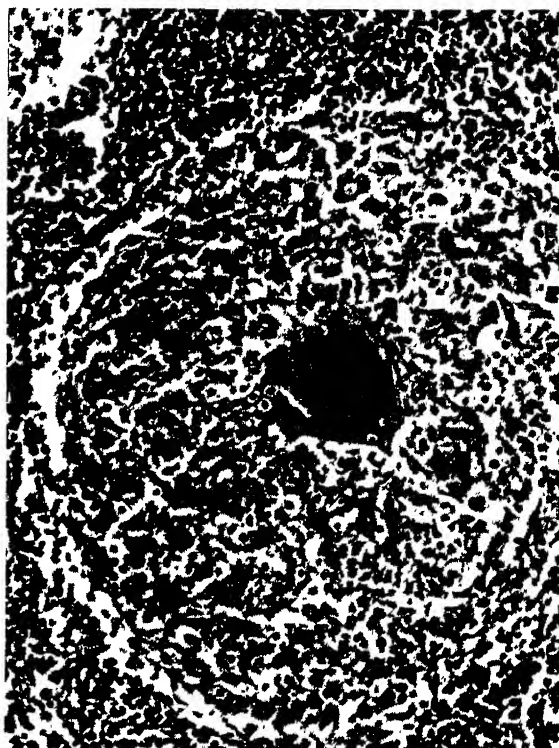


Fig. 8.13 An early tubercle, consisting mainly of epithelioid cells, some of which have fused to form a Langhans' giant cell. Lymphocytes are scattered among the epithelioid cells and are numerous around the periphery. $\times 174$.



Fig. 8.14 A more advanced tubercle with three giant cells and early necrosis among the most centrally-placed epithelioid cells. $\times 150$.

slightly granular material, which may also contain fibrin from vascular exudation. The tubercle thus comes to consist of a necrotic centre, surrounded by epithelioid and sometimes giant cells (Fig. 8.15), with a peripheral aggregation of small lymphocytes. The necrotic material is creamy-white, and resembles cream cheese in appearance and consistence—hence the terms **caseation** and **caseous material**.

The initial accumulation of macrophages and phagocytosis of tubercle bacilli occur before there is any immune response and are seen also in the reaction to particles of various non-antigenic foreign materials. Lymphocytic infiltration, however, follows the development of cell-mediated immunity and some of the cells are specifically primed T-cells which, by releasing various lymphokines (p. 159) contribute to the arrival of more macrophages by chemotaxis, and to their arrest around the tubercle bacilli by migration-inhibition factor: macrophage-activating factor may transform the macrophages to epithelioid cells, and may mediate the destruction of phagocytosed bacilli. The T-cell cytotoxic factor presumably accounts for the necrosis of macrophages at the centre of the lesion, although the tubercle follicle is avascular and ischaemia may also be important. It

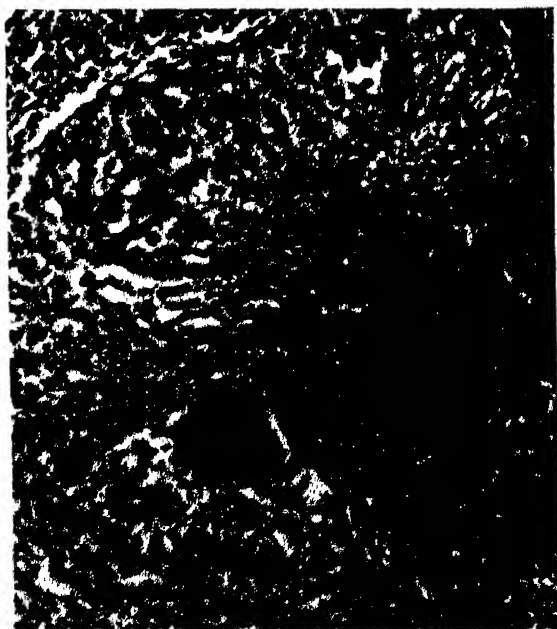


Fig. 8.15 Part of a tuberculous lesion with central caseation (*right*) and surrounding epithelioid cells with giant-cell formation. Note the structureless appearance of the caseous material. $\times 150$.

has been shown in animal experiments that many of the lymphocytes in the tuberculous lesion are not specifically primed cells resulting from the cell-mediated immune response (p. 157): they may be attracted to the site of infection by a chemotactic lymphokine, but their significance in the lesion is obscure.

The further course of the infection depends on several factors, including the infecting dose and virulence of the organism and also the degree of resistance of the host. What determines virulence in strains of tubercle bacilli is not understood, but the so-called virulent strains are those which are capable of relatively rapid multiplication *in vivo*. If bacterial multiplication is checked, tubercles are replaced by fibrous tissue. If the bacteria continue to multiply in the lesions, they may escape and gain a foothold in the surrounding tissues, with further tubercle formation. A cluster of tubercles may thus arise, and as these enlarge, they become confluent, and the central areas of caseous necrosis eventually unite to give a large caseous patch with tubercles around the periphery. Such lesions may reach several centimetres in diameter. When they arise in the lungs, they seldom reach this size without involving the wall of a bronchus, and the caseous material is then discharged, leaving a tuberculous cavity (Fig. 8.16). In other tissues, and particularly in the kidneys and in lesions of bone



Fig. 8.16 Formation of a tuberculous cavity. The lesion has ulcerated into a bronchus and the caseous material is discharging. $\times 4$.

extending into the surrounding soft tissues, caseous material may be invaded by **neutrophil polymorphs**, with resultant **liquefaction** ('tuberculous pus'). Such a lesion used to be called a **cold abscess**, because it is not accompanied by the acute inflammatory features of a pyogenic abscess. The softened caseous material may track through the tissues and may eventually reach a surface and discharge.

Some tuberculous lesions present more acute features than those described above. For example, rapid dissemination may occur by the air passages throughout the lung, resulting in multiple scattered lesions. Microscopy then shows extensive filling of the alveoli with large rounded macrophages (Fig. 8.17); these cells rapidly undergo fatty change and necrosis, and the lesions enlarge and coalesce with little or no attempt at healing. If it infects the subarachnoid space, usually by way of the bloodstream, the tubercle bacillus multiplies rapidly in the cerebrospinal fluid and the meningitis is of acute exudative inflammatory type with deposition of fibrin and accumulation initially of neutrophil polymorphs and later of macrophages and lymphocytes. Tubercles are usually poorly formed, and involvement of the walls of arteries and veins lying in the subarachnoid space may cause severe narrowing of their lumina by endarteritis (Fig. 21.33, p. 751) or occlusion by thrombosis. The lesions which result from infection of the pleural and peritoneal cavities are also commonly exudative, with a serous or serofibrinous exudate, and when the pericardium is involved the exudate may be rich in fibrin and is often haemorrhagic, presumably as a result of the mechanical effect of the heart beat.

Primary and reinfection tuberculosis

Infection of an individual who has not been previously infected or immunised gives rise to the **primary lesion** at the portal of entry in the lung, tonsil or small intestine. This usually remains small, and commonly heals without becoming detectable. Early spread of bacteria to the regional lymph nodes is, however, the rule, and their rapid multiplication may occur in the affected nodes, i.e. at the root of the lung (Fig. 16.29, p. 476), in the neck or in the mesentery, depending on the site of the primary lesion. The combination of the primary lesion and

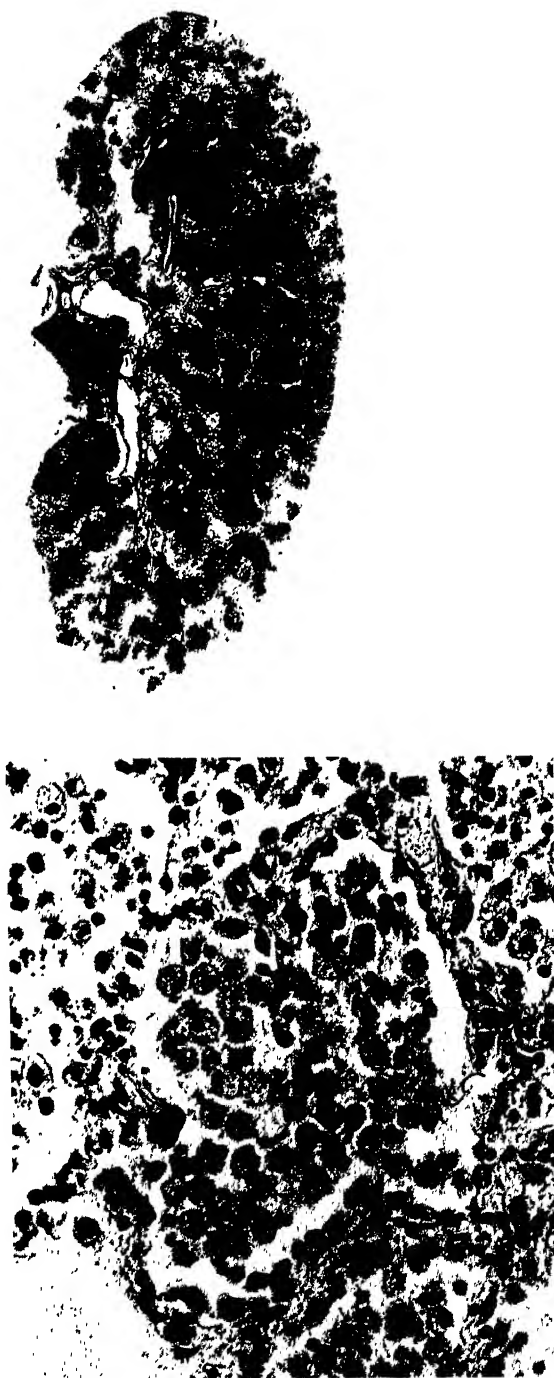


Fig. 8.17 Acute tuberculous bronchopneumonia. *Above*, Primary pulmonary tuberculosis in a child. The infection has spread by the bronchi and has caused widespread lesions which are becoming confluent and have undergone central caseation (dark areas). *Below*, the lesions consist initially of accumulations of macrophages and lymphocytes in the alveoli. The macrophages undergo fatty change and appear 'foamy': necrosis then supervenes. $\times 370$.

enlarged, caseous regional lymph nodes is called the **primary complex**.

Reinfection or **chronic tuberculosis** results from infection of an individual who has overcome a primary infection or been immunised by BCG. The reinfection lesion is usually in the apex of one or other lung and may extend to give a large local lesion with one or more cavities (Fig. 8.18). There is usually little or no involvement of the local lymph nodes. Reinfection lesions occur also in the tonsils, small intestine, pharynx and skin, again without much involvement of the regional nodes, but these are relatively uncommon sites. Individuals with reinfection tuberculosis of the lungs may, however, develop lesions in the larynx, mouth and intestines as a result of endogenous infection by coughing up and swallowing sputum containing tubercle bacilli (Fig. 19.59, p. 633). These metastatic lesions resemble those of reinfection tuberculosis in spreading locally with minimal or no involvement of the local nodes.

The differences between primary and reinfection tuberculosis appear to depend mainly on the spread of the bacilli to the local lymph nodes and their multiplication there in the early

stages of the primary infection, before the development of a high level of cell mediated immunity.

Although the term 'reinfection tuberculosis' is commonly used for chronic tuberculosis occurring usually in adults, it may well be that, in some cases, tubercle bacilli have persisted in a healed primary lesion and eventually multiply and produce the 'reinfection'.

The spread of infection within the body is discussed below, but more detailed accounts of the resulting lesions are given in the chapters on regional pathology, e.g. pulmonary tuberculosis, pp. 475 *et seq.*

Amyloid disease (p. 269) is an important complication of chronic tuberculosis.

Spread of infection

Tuberculous infection is very prone to spread by lymphatics and to produce lesions in lymph nodes. This occurs especially in the early stages of the primary infection, with resulting involvement of the draining lymph nodes, and infection may spread from these to adjacent nodes or groups of nodes, e.g. in the mediastinum. In chronic (i.e. reinfection) tuberculosis, lymphatic spread is usually localised to the tissue immediately around the lesions, the draining lymph nodes seldom being severely involved. This limitation of lymphatic spread is probably attributable to the modified behaviour of macrophages which results from delayed hypersensitivity. There is experimental evidence that lymphatic spread results from ingestion and transport of *Myco. tuberculosis* by macrophages, and the T-cell factors which convert macrophages into 'killer' cells and interfere with their migration from the lesion are likely to impede such spread.

Spread also occurs by the bloodstream. This is seen notably in **acute miliary tuberculosis**, in which large numbers of bacteria enter the blood and give rise to multiple scattered tubercles in the various organs. The condition arises most commonly in primary tuberculosis and is due usually to involvement of a vein by the large caseating lymph node lesions of the primary complex—in most cases the pulmonary hilar nodes: the caseating process extends into the wall of an adjacent vein, usually one of the pulmonary veins, and caseous material containing large numbers of mycobacteria



Fig. 8.18 Apical part of the lung showing chronic (reinfection) tuberculosis. The infection has extended to form coalescing lesions with central caseation and peripheral fibrosis. The caseous material in the larger lesions has discharged via the bronchi, leaving several cavities with fibrous walls. $\times 0.7$.

is then discharged into the circulation. The resulting lesions are particularly numerous in the liver, kidneys and spleen. They consist of tubercles of fairly uniform size, and without specific therapy death usually results from tuberculous meningitis after about a month, at which time the tubercles are of approx. 1–2 mm diameter (Fig. 16.31, p. 479): they are rather poorly developed, often without giant cells, but with central necrosis (Fig. 8.19) and are termed **miliary tubercles** (latin *milium*—millet seed). In some instances, the bacteria escape into a systemic vein, either directly or by involvement of the thoracic duct, and as a result the number of miliary lesions in the lungs far exceeds those in other organs. When a relatively small number of tubercle bacilli gain entrance to the bloodstream, few tubercles are produced in the various organs, and since the patient may survive much longer than is the case in untreated acute miliary tuberculosis, the lesions may become larger. One or more large metastatic lesions may also occur, for example in the bones, joints, kidneys, epididymes or fallopian tubes, and less commonly in the brain. Although in heavily infected communities blood-borne lesions arise most commonly as a complication of

the primary tuberculous complex in young children, in countries where the disease has been largely eradicated they are uncommon and are now seen mostly in older patients with reinfection tuberculosis, particularly in advanced cases. Haematogenous lesions are also observed when tuberculosis is complicated by other debilitating diseases, or as a result of corticosteroid or other immuno-suppressive therapy.

Spread of tuberculous infection occurs also along hollow viscera and in body cavities. In the lungs, spread by the bronchi is of great importance. Mycobacteria coughed up in sputum may settle and give rise to lesions in the larynx and intestine. Spread may occur from the fallopian tubes to the endometrium, and from the kidney to the urinary tract, and dissemination may occur within the pleural, pericardial and peritoneal cavities and, in tuberculous meningitis, within the subarchnoid space and ventricles of the brain.

Healing of tuberculous lesions

The healing of tubercles or larger tuberculous lesions is dependent on the elimination or

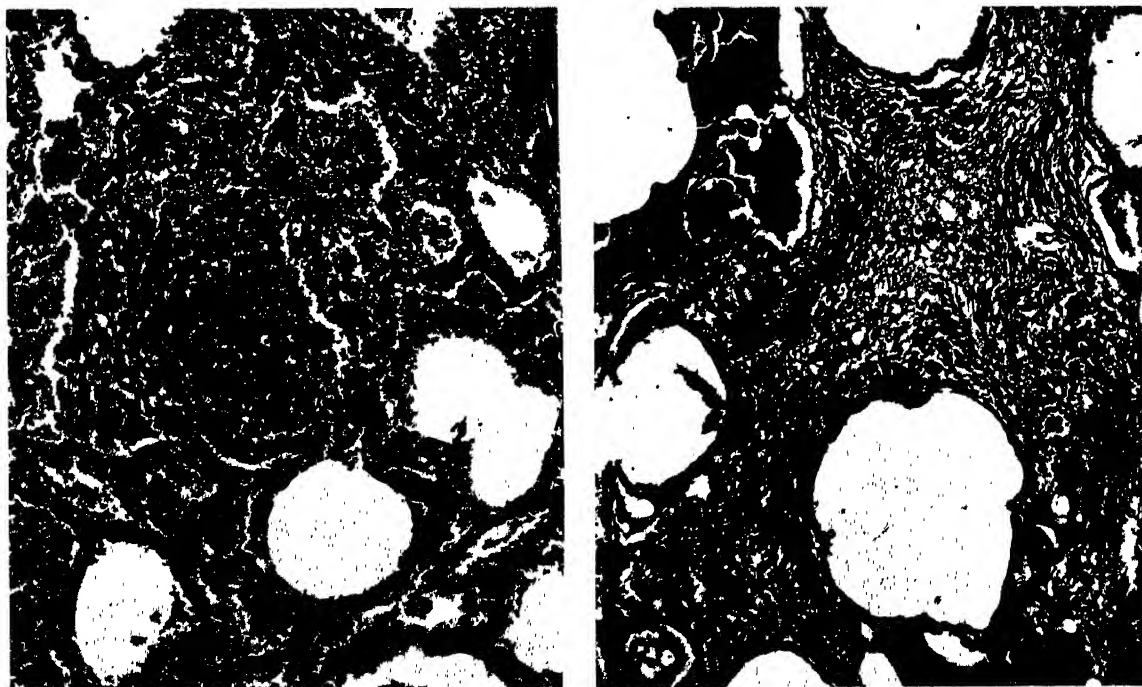


Fig. 8.19 On the left is shown a miliary tubercle of lung with central caseation and acute exudate in the surrounding alveoli from a patient with untreated miliary tuberculosis; on the right a fibrous scar containing a few lymphocytes, the remains of a miliary tubercle after specific chemotherapy. $\times 100$.

reduction in numbers of mycobacteria. Healing is brought about by formation, around the lesions, of reticulin fibres, progressing to more dense fibrosis. If caseation is slight or absent, as in early tubercles, the whole lesion may be gradually replaced by fibrous tissue, leaving a scar (Fig. 8.19), but extensive patches of caseation usually persist and become encapsulated in fibrous tissue. Slow progressive deposition of calcium salts commonly occurs in the caseous material, which eventually may become stony hard and clearly visible in radiographs: in some instances the calcified material may be replaced by bone, which may even develop spaces containing haemopoietic marrow. The course of the disease depends on the balance between bacterial multiplication, with extension and caseation of the lesions, and the reactive processes involved in killing the bacteria, preventing their spread, and promoting a fibroblastic reaction.

The effects of drugs. The course and prognosis of tuberculosis have been radically changed by effective specific chemotherapy, which has also greatly modified the appearance of tuberculous lesions. When healing occurs 'naturally', i.e. without specific chemotherapy, large caseous lesions become walled off first by cellular tubercles and then by new-formed fibrous tissue, which penetrates the outer zone of tubercles and finally encapsulates the central caseous mass. Dense fibrosis and calcification complete the process: there is little resolution. Effective drug therapy is accompanied successively by resolution of the surrounding exudative lesions, increased vascularity, reversion of the epithelioid cells to foamy macrophages, formation of granulation tissue, absorption of necrotic and caseous material, and finally by healing with the production of minimal amounts of fibrous tissue. Combined therapy thus strikingly modifies the outcome; early lesions may clear up almost completely without residual effects and chronic caseous and fibrotic pulmonary lesions with excavation are transformed to smooth-walled cavities, which may become lined by epithelium.

Because of the spontaneous occurrence of mutant tubercle bacilli resistant to one or more drugs, it is now accepted practice to administer a combination of three drugs, usually isonicotinic acid hydrazide (isoniazid), rifampicin and ethambutol. Because the tubercle bacillus mul-

tiplies relatively slowly, it may take some months for the selective advantage conferred by drug therapy on a resistant mutant to be reflected in clinical deterioration, and repeated bacteriological examinations, e.g. of sputum, are therefore important during treatment.

Leprosy

It is estimated that there are some 10 million people with leprosy throughout the world and approximately 100 000 new cases are registered annually. The disease now occurs mainly in those tropical and sub-tropical countries with poor living standards, although it was formerly quite common in Europe and North America. The causal organism, *Mycobacterium leprae*, was the first bacterial pathogen to be seen and described in a human disease. It is an acid-fast bacillus, demonstrable by a modified Ziehl-Neelsen staining technique: it has not yet been grown in culture and does not cause natural disease in species other than man, although it causes a local lesion when injected into the mouse foot-pad and spreading infection in animals in which cell-mediated immunity has been depressed by neonatal thymectomy or anti-lymphocyte serum. *Infectivity is low, and although the bacillus is often present in large numbers in the nasal and oral secretions of patients, only a small percentage of long-term close contacts develop the disease.*

The leprosy bacillus appears to be highly temperature-dependent, for it produces lesions mainly in colder parts of the body, namely the skin, especially of the nose, lobes of the ears and extremities, the anterior part of the eye, the nasopharynx, mouth and upper respiratory tract, superficial lymph nodes and testes. It also has a predilection for nerves; it always involves the small nerve twigs in the skin and often larger superficial nerves.

Host resistance is dependent mainly on the cell-mediated response to *Myco. leprae*. In **lepromatous leprosy** this is depressed, and the lesions progress relentlessly and contain huge numbers of bacilli. A strong response is associated with **tuberculoid leprosy**, in which the lesions contain relatively few bacilli, and may become stationary or subside.

Lepromatous leprosy. The lesions consist of aggregates of macrophages (Fig. 8.20), containing huge numbers of lepra bacilli lying parallel



Fig. 8.20 Leprosy. *Above*, part of a skin nodule in lepromatous leprosy. The lesion is a macrophage granuloma: lymphocytes are scanty and necrosis has not occurred. $\times 60$.

Middle, part of the same lesion, stained by Triff's method for acid-fast bacilli: the macrophages contain numerous *Myco. leprae*. $\times 650$.

in bundles or aggregated to form large acid-fast masses (globi). In the absence of cell-mediated immunity, the bacilli are not destroyed, and multiply within the macrophages which have ingested them. Like tubercle bacilli, they appear to be non-toxicogenic and macrophages containing large numbers of organisms show little evidence of injury apart from fatty change. Lymphocytes are scanty or absent, and tissue injury is probably due mainly to pressure. However, the skin lesions become very extensive and may ulcerate and become secondarily infected with various bacteria. Similar changes are seen in the oral, nasal and upper respiratory lesions. In spite of extensive involvement of superficial nerves, anaesthesia, trophic changes and paralysis are often late features of the disease. The internal organs are rarely seriously involved, although small clusters of macrophages containing lepra bacilli may occur in the liver, spleen, etc.

Tuberculoid leprosy. The lesions resemble those of tuberculosis, consisting of tubercle-like follicles (Fig. 8.20) and more extensive infiltrates of epithelioid and Langhans' giant-cells (both of which are modified macrophages) and lymphocytes. Necrosis occurs but is inconspicuous and leprosy bacilli are often difficult to find. Nerve injury due to scarring occurs early, and patches of hypopigmentation and sensory loss are often presenting features. The internal organs are seldom involved.

Forms of leprosy intermediate between lepromatous and tuberculoid are commonly encountered. In some instances, lesions of both types may be present (**borderline or dimorphous leprosy**), and such cases may progress in either direction. Drug therapy helps to convert lepromatous to tuberculoid leprosy, and to cure the latter.

Immunological reactions in leprosy. Intradermal injection of lepromin, an antigenic extract of *Myco. leprae*, elicits a delayed hypersensitivity reaction, resembling the tuberculin reaction, in individuals with tuberculoid leprosy: the reaction occurs also in individuals with cell-mediated immunity to the tubercle bacillus and

Below, skin lesions in tuberculoid leprosy, composed of epithelioid cells, giant cells and lymphocytes. Such lesions contain few bacilli. $\times 64$. (The late Professor J. A. Milne.)

other mycobacteria. A later granulomatous lesion (the Mitsuda reaction) may develop within 4 weeks: this also is not diagnostic of leprosy. The test is usually negative in lepromatous leprosy, and becomes positive if the patient converts, spontaneously or as a result of drug therapy, to tuberculoid leprosy. These observations are supported by parallel results of *in-vitro* tests for cell-mediated immunity, and together with the morphological features of the disease indicate that *tuberculoid lesions are modified by a delayed hypersensitivity reaction to the lepra bacilli, whereas lepromatous lesions represent the growth of the bacilli in the absence of delayed hypersensitivity.*

Failure of cell-mediated immunity in lepromatous leprosy is not fully explained. The lymph nodes show lymphocyte depletion of the T-dependent zones, i.e. paracortex, and diminution of T lymphocytes in the blood has been reported. A plasma factor which depresses T lymphocyte function has also been described.

There is usually a high titre of antibody to *Myc. leprae* in the serum of patients with lepromatous leprosy, but this seems to afford little protection: drug therapy results in destruction of large numbers of bacilli, and release of antigen may then give rise to an acute Arthus reaction in the lesions (*erythema nodosum leprosum*) or generalised immune complex disease with glomerulonephritis (pp. 155–6). Another complication is amyloid disease.

Sarcoidosis

This disease, which is of unknown causation, is characterised by multiple granulomatous lesions, and may affect lymph nodes, lungs, skin, spleen, eyes, salivary glands, liver and bones, particularly of the hands and feet. It is of world-wide distribution, but with great geographical variation in incidence. It occurs over a wide age range, but most commonly in young adults, and is much more common in negroes than whites in the U.S.A., and in immigrants than natives in Great Britain. The highest reported incidence is in Sweden.

The disease most commonly gives rise to enlarged mediastinal and pulmonary hilar lymph nodes, often without symptoms, but sometimes accompanied by fever. Other groups of lymph nodes are often affected and minute lesions in the lungs may present an x-ray picture resem-

bling that of miliary tuberculosis. Sarcoid lesions also occur in the skin and occasionally erythema induratum (p. 1065) develops and may be the presenting clinical feature. Microscopically, the sarcoid lesions consist of tubercle-like follicles composed of epithelioid cells with occasional giant cells and scanty peripheral lymphocytes (Fig. 8.21 and Fig. 18.10, p. 574). The giant cells may contain curious calcium-rich star-shaped or conchoid inclusions (asteroid or Schaumann bodies). Unlike tuberculosis, the lesions do not undergo caseation although there may be a little central necrosis.



Fig. 8.21 Sarcoidosis of skin. The lesions consist of aggregates of epithelioid cells with relatively few lymphocytes. In contrast to tuberculosis, there is little or no necrosis. $\times 80$.

The course of the disease is unpredictable: it may be acute or chronic, and temporary or permanent remission may occur spontaneously. It can cause blindness by involving the uveal tract, and is occasionally fatal, usually as a result of fibrosis of the pulmonary lesions with consequent right ventricular heart failure, or as a result of intercurrent infections. Hypercalcaemia may develop with consequent renal damage.

Diagnosis. Non-caseating epithelioid-cell granulomas, with or without giant-cell inclusions, are not diagnostic of sarcoidosis: they occur in various conditions including tuberculosis, various fungal infections, syphilis, brucellosis and berylliosis. The diagnosis of sarcoidosis thus

depends also on the clinical features, the distribution of the lesions, and on excluding the above possibilities. Sarcoid-like follicles are occasionally found incidentally in surgical and necropsy material and are of unknown significance.

Intradermal injection of a sterile suspension prepared from sarcoid lesions (**Kveim test**) leads to the development of a lesion becoming maximal in about six weeks and having the histological features of sarcoidosis. The test is positive in most cases of sarcoidosis, but conflicting results have been reported in some other conditions, notably Crohn's disease (p. 620). During the course of sarcoidosis, tuberculin tests are negative in most cases even when these are known to have been previously positive and cell-mediated immunity in general is impaired. A fall in circulating T lymphocytes and a depressed response of T cells to PHA and other mitogens (p. 172) have also been reported. The capacity to produce antibodies, however, is normal or even increased.

Aetiology. The significance of these immunological features is obscure. The sarcoid lesion, consisting of epithelioid cells and lymphocytes, is itself suggestive of a delayed hypersensitivity reaction, although no exogenous antigen has been shown to be involved.

Subsequent development of tuberculosis has been observed in some patients, but sarcoidosis seems unlikely to be a modified form of tuberculosis, because depression of delayed hypersensitivity (as in sarcoidosis) would be expected to be associated with a florid form of tuberculosis. Also, the condition is not aggravated, and is sometimes improved, by administration of steroids. Various other aetiological factors have been suggested, but with little good supporting evidence.

Syphilis

Historical note

It is generally believed that syphilis was introduced into Europe on the return of the Spanish sailors of Columbus from America and that by the end of 1494 it had spread throughout Spain and along the Mediterranean coast into Italy. Within a century it had become widespread throughout Europe, having been carried everywhere by the mercenary troops returning to their own countries after the Siege of Naples (1495). At this time syphilis was clearly re-

cognised as a new disease and its manifestations became so well known that Shakespeare was able to give a remarkably accurate (although anachronistic) account of them in *Timon of Athens* (Act IV, Scene 3). Absence of syphilis from the Old World is supported by the complete lack of evidence of the disease in skeletal remains dating back from 1494, whereas bones found in ancient tombs in Central America bear clear indications of the disease. The name comes from a poem composed in 1530 by Girolamo Frascatoro, a Verona physician, in which Syphilis, a swineherd, offended Apollo, who inflicted him with the disease.

General features

Formerly common, syphilis is now relatively infrequent in Western Europe: recent reports show some increase, particularly among homosexuals, but the rise is much less than for gonorrhoea. Syphilis is an important **venereal disease**, i.e. it is usually contracted by coitus and the primary lesion then develops on the external genitals. Rarely, extragenital infection occurs on the lip, tongue or breast and also on the fingers from handling infective lesions. The causal agent is a small motile spiral micro-organism or spirochaete, *Treponema pallidum*. It dies rapidly on drying and even if kept moist does not survive for long outside the body. Accordingly, infection is usually by direct contact, the presence of a minute abrasion or crack in the skin apparently facilitating invasion. The disease has a distinct *incubation period*, followed by a *primary lesion* and then a *febrile secondary stage* with skin eruptions, and this is sometimes followed by a *tertiary stage*, with localised lesions, and by a late stage of *neurosyphilis*. It is convenient to give a general survey of the course of the untreated disease at this point. The special features of the individual lesions will be considered in the appropriate systematic chapters.

'Stages' of syphilis

The primary sore. The primary sore or chancre (Fig. 8.22) appears usually on the external genitals, after an incubation period of 2–12 (usually 3–4) weeks, as a small, slowly growing, hard, pale brownish-red, usually painless nodule. The centre ulcerates and there may be some exudate which, in a skin lesion, is usually scanty and forms a crust. When the lesion is on



Fig. 8.22 Primary syphilitic chancre of penis. The lesion is seen as a swelling with (in this instance) central necrosis and ulceration: it is situated in the coronal sulcus and involves the reflection of the prepuce.

a mucous surface and the part is not kept clean, there may be more extensive ulceration, and various organisms, sometimes including other spirochaetes, are present in addition to *Tr. pallidum*. The ulcer persists for some weeks, during which the inguinal lymph nodes, usually on both sides, become somewhat enlarged and hard. *Treponema pallidum* is often detectable in the exudate of the ulcerated chancre, either by dark-ground microscopy or by fluorescence microscopy, using fluorescein-labelled antibody: if this fails, it may be demonstrable in fluid withdrawn by puncture of the enlarged lymph nodes. *Dissemination by the blood takes place before the appearance of the primary lesion*, and syphilis has been accidentally transmitted by transfusion of blood withdrawn before the primary lesion had appeared in the donor.

The primary chancre subsides spontaneously after a few weeks, leaving a slight scar. In a significant proportion of cases, it does not develop or passes unnoticed.

Secondary lesions appear at a variable interval, usually from 2–3 months after infection; they include multiple symmetrical lesions of the skin and squamous mucous membranes. The skin rash may be macular, papular or pustular, the palms of the hands and soles of the feet being commonly involved. Lesions of the hair

follicles in the scalp lead to loss of the hair—*alopecia*. In the vulva, anus and perineum, flat raised papules sometimes develop—*condylomata lata*—and are intensely infective: they must not be confused with *condylomata acuminata*, the so-called venereal warts, which are of viral nature. The buccal and pharyngeal mucosa shows white, shining patches caused by thickening of the keratinised layer, and these break down, giving ‘*snail-track ulcers*’. General slight enlargement of lymph nodes is also common and is most easily detected in the superficial nodes. The secondary lesions are usually accompanied by fever, anaemia and general malaise. After some months all these features disappear spontaneously and the disease becomes latent.

Tertiary lesions appear irregularly, especially in the internal organs, skin and mucous membranes; they are few in number but usually much larger than the primary and secondary lesions, and lead to serious and permanent damage. They rarely appear within the first few years, and sometimes only after many years. Tertiary lesions are characterised by diffuse chronic inflammation, often with central necrosis, and extensive formation of granulation tissue. If necrosis is present, the lesion is termed a **gumma**. The central necrotic portion is dull yellowish, firm and rubbery; this is surrounded by a more translucent capsule of young connective tissue which has often a very irregular outline (Fig. 25.8, p. 996). Tertiary lesions may occur in any tissue, but especially in the liver, testes and bones. They cause extensive destruction, e.g. in the nasal bones with loss of the bridge of the nose and perforation of the palate, ulceration and destruction of the larynx, creeping ulcers in the skin, etc. Of special importance are the cardiovascular lesions. All tertiary lesions tend to heal eventually, and much distortion of the organs and interference with function may result from scarring.

Neurosyphilis. Lastly, in a small proportion of cases there occur two important nervous diseases, *tabes dorsalis* and *general paralysis*. They are due to the actual presence of the spirochaetes in the central nervous system.

Microscopic appearances

The main feature of the early **chancre** is heavy cellular infiltration of the dermis (Fig. 8.23)



Fig. 8.23 Primary syphilitic chancre, showing the heavy cellular infiltration of the dermis. Most of the infiltrating cells (not readily identified at this magnification) are lymphocytes and plasma cells. $\times 120$.

with lymphocytes, plasma cells and occasional macrophages, which are mainly responsible for the hardness and swelling. At the periphery, the infiltrating cells lie mainly around the small vessels (periarteritis). Later, ulceration occurs with exudative inflammation and formation of granulation tissue. *The histological features are not diagnostic without the demonstration of *Tr. pallidum*, which requires special staining techniques.** After a time the cellular infiltration gradually diminishes and only a little thickening of the fibrous stroma remains. There is usually little or no residual scarring unless there has been much ulceration.

In the **secondary lesions** in the skin and mucous membranes the main changes are vascular engorgement and infiltration, mainly of plasma cells, but also lymphocytes and macrophages (Fig. 8.24). Cellular infiltration occurs also around and into the hair follicles, and the hairs may fall out. All these disseminated lesions of the skin and mucous membranes usually subside naturally, i.e. without specific therapy, and without scarring.

The **gumma** of the **tertiary stage** consists of



Fig. 8.24 Papular syphilitic rash, showing abundant cellular infiltration of the corium. Note also the hyperkeratosis. $\times 115$.

parenchymal necrosis, surrounded by a layer of connective tissue infiltrated with lymphocytes and plasma cells (Fig. 8.25). Eventual healing is accompanied by shrinkage, considerable scarring and distortion. Another common type of lesion is chronic interstitial inflammation or fibrosis, often spreading extensively, and sometimes containing foci of gummatous necrosis. In the necrotic tissue, the structural outlines may be preserved for a long time, the cells not having the same tendency to fuse into amorphous material as is seen in caseous tuberculosis. Giant cells may be present in the granulation

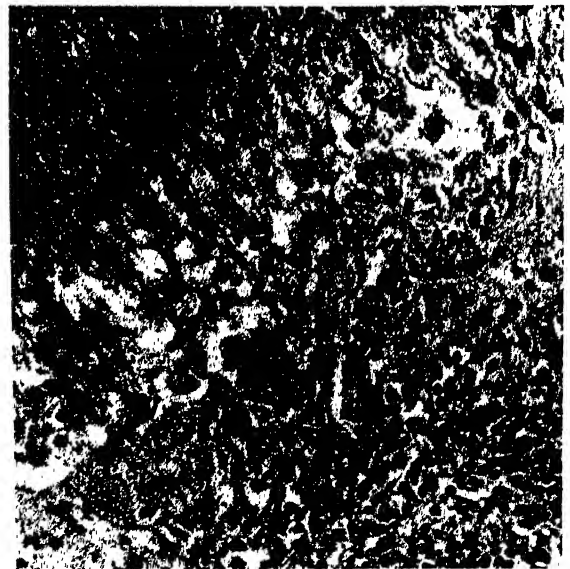


Fig. 8.25 Section of part of a gumma, showing the necrotic centre (*upper left*), bounded by connective tissue heavily infiltrated with lymphocytes. $\times 250$.

* *Tr. pallidum* is argyrophilic (p. 648) and is thus stained by Levaditi's and similar methods based on deposition of silver.

tissue at the periphery, but they are usually smaller than in tuberculosis, and there are no well-formed follicles. Nevertheless the histological diagnosis between the two diseases may be difficult. The important **vascular lesions** of syphilis are described later; those in the larger arteries are due to the presence of spirochaetes in the adventitial sheath and media: they give rise to cellular infiltrations like those described above, and some medial necrosis may follow.

The number of spirochaetes in gummatous lesions is small, and the necrosis seems to be due either to ischaemia resulting from endarteritis of small vessels, or to a hypersensitivity reaction, possibly of delayed type.

A non-venereal form, termed *endemic syphilis* occurs in children in parts of Africa and India, and a similar condition, *bejel* affects young children in Arab countries.

Congenital syphilis

The first pregnancy after untreated infection is likely to terminate prematurely with a mace-rated fetus, in the tissues of which spirochaetes are abundant. The parenchymatous organs show diffuse proliferation of fibroblasts with minute foci of necrosis—miliary gummas—and there is severe damage to the liver, lungs, pancreas, etc. In subsequent pregnancies the effects are progressively less severe. The next child may be born alive with lesions of congenital syphilis, including a papular rash around mouth and nose, on the buttocks, palms of hands and soles of feet. Disease of the nasal bones and mucosa leads to 'snuffles' and interference with feeding. Syphilitic hepatitis with jaundice, splenomegaly, and lesions in the bones are also common. Later a characteristic deformity appears in the incisor teeth, which are peg-shaped with notched edges (Hutchinson's teeth) and there is also pitting of the first permanent molars. Still later, neurosyphilis may develop and also interstitial keratitis causing corneal opacity and blindness. Pregnancy has a curiously ameliorating effect on syphilitic lesions in the mother, who may appear healthy in spite of producing syphilitic offspring.

Immunology of syphilis

At least three distinct antibodies develop in syphilis, and their detection in the serum is of

considerable diagnostic value. The older tests are based on the detection of antibody reactive with the diphosphatidylglycerol component of phospholipids of mitochondrial membranes. Alcoholic extract of beef heart muscle ('*cardiolipin*') is employed as antigen, but extracts of various normal animal and human tissues may also be used. Antibody is demonstrable by various precipitation (flocculation) techniques, e.g. the Kahn, Kline or VDRL (venereal disease research laboratory) tests, or by the Wassermann test which is based on complement fixation (p. 112). These are the so-called **standard tests for syphilis**: they are useful screening tests, antibody being detectable from an early stage, but are not specific for syphilis, **false positive reactions** occurring in various conditions, including many acute infections, malaria, infectious mononucleosis, mycoplasmal pneumonia, trypanosomiasis, leprosy and systemic lupus erythematosus. The tests are also positive occasionally in apparently normal individuals, particularly during pregnancy. It is not understood why antibody to diphosphatidylglycerol develops in syphilis: it fulfils the criteria of an auto-antibody (p. 161), and the false positive reactions may result from auto-immunisation as a result of tissue destruction from causes other than syphilis, with release of cellular constituents.

A second antibody, which reacts with group antigen common to various species of treponemes, may be detected by a complement-fixation test, using as antigen non-pathogenic treponemes which grow readily in culture; this is also used as a screening test but is not specific for syphilis.

Confirmatory tests for syphilis depend on the demonstration of antibody specific for *Tr. pallidum* by (a) the treponemal immobilisation test in which the patient's serum is added to a suspension of living *Tr. pallidum* and antibody is indicated by immobilisation of the treponemes, or (b) the fluorescent antibody technique in which binding of antibody to *Tr. pallidum* is demonstrated by means of fluorescein-labelled anti-immunoglobulin (p. 111).

Antibody tests usually become positive a week or so after the appearance of the primary lesion: they are virtually always positive in the secondary stage, following which the percentage of positives falls. In neurosyphilis, antibody is more likely to be detected in the cerebrospinal fluid than in the serum.

Following cure, the antibody tests become negative, although the specific treponemal antibodies may persist for some years.

The protective role of specific immunity in syphilis is suggested by the overwhelming infection which sometimes occurs in the immunologically immature fetus. Some of the features of secondary syphilis, including the widespread skin lesions and occurrence of arthralgia and occasionally glomerulonephritis, are probably due to the union of large amounts of treponemal antigen with circulating antibody, to form immune complexes (pp. 155–6). In both congenital and secondary syphilis, lymphocyte depletion in the T-dependent areas of the spleen and lymph nodes, and depressed T-cell function, have been reported.

Other treponemal diseases

Two other diseases caused by treponemes occur in tropical countries. One is **yaws**, which resembles syphilis but is non-venereal and rarely causes cardiovascular or neurological disease. The causal agent, *Tr. pertenuae* cannot be distinguished from *Tr. pallidum* and infection with either confers immunity to both: the distinction from syphilis is based on clinical features. The second condition is **pinta**, another non-venereal chronic disease somewhat resembling syphilis: it is caused by *Tr. carateum* which does not confer immunity to syphilis.

Other pathogenic spirochaetes

These include the **Borreliae**, which are transmitted by lice and ticks, and cause **relapsing fever** (p. 696) and the **Leptospirae** which infest rodents, etc., and cause febrile illnesses, the best known being **Weil's disease** (p. 695).

Actinomycosis

This disease is produced by organisms which are normal commensals in the mouth and gut and only occasionally invade the tissues to produce infection. The actinomyces are branching bacteria which grow in the tissues to produce characteristic radiate colonies, sometimes visible macroscopically. In man, the micro-aerophilic *Actinomyces israelii* is the chief pathogen, but occasionally aerobic organisms—*Nocardia*—are involved, and also other species,

which grow more diffusely. In bovines, in which actinomycosis due to *Actino. bovis* is common, the lesions are localised and are large granulomatous masses which occur especially in and around the jaw. In man the disease usually affects children and young adults, more often males than females, and agricultural workers appear to be particularly at risk. The lesions are of a more suppurative type, and in about 70 per cent of cases are in the region of the mouth or jaws, the parasite gaining entrance commonly from a tooth socket following extraction or from a carious tooth. In 15 per cent the infection is in the appendix or caecal region, from which spread by the blood stream to the liver may occur; in about 10 per cent the initial lesion is in the lung and in 5 per cent it is subcutaneous. The lesion is usually a chronic suppurative one, with formation of multiple abscesses, each containing one or more colonies of the organism—the so-called honeycomb abscess. Fibrous septa between the abscesses are lined by granulation tissue which contains many foamy cells—macrophages laden with lipid—giving the lining of each abscess a yellowish colour. In the centre is pus containing actinomyces colonies (Fig. 8.26), which are sometimes visible by naked eye as small yellow or grey, gritty granules ('sulphur granules').

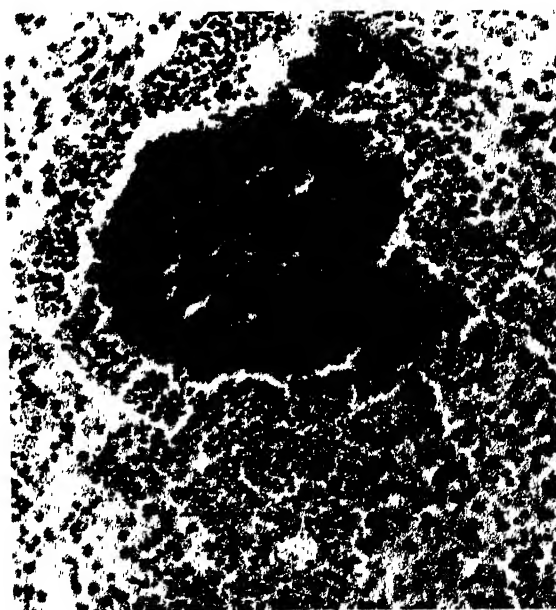


Fig. 8.26 Actinomycosis. A colony of *Actinomyces israelii* in a small abscess, the wall of which consists of granulation tissue heavily infiltrated with lipid-laden (foamy) macrophages. $\times 190$.

Lesions of the face and neck, originating about the jaw, may produce much granulation tissue in which many small foci of suppuration persist and discharge through the skin, resulting in multiple sinuses. The infection spreads directly

through the tissues but does not usually involve the regional lymph nodes; if untreated it tends to invade the bloodstream, giving rise to pyaemia with secondary abscesses in the liver, lungs and other organs.

Other Types of Infection

Rickettsial infections

The rickettsiae are micro-organisms of various shapes, smaller than bacteria but resembling them in their structural and metabolic features, including presence of a cell wall. They are obligatory intracellular parasites and infect many species including arthropods, birds and mammals. Several species of rickettsiae cause disease in man: in most instances they enter the body by the bites of infected ticks or mites, or from infected louse or flea faeces being scratched into the skin. The organisms enter and multiply in the endothelium of the capillaries and other small blood vessels; they are at first localised to the site of infection, but blood dissemination occurs during the incubation period and endothelial involvement then becomes widespread. Capillary obstruction from endothelial swelling or thrombosis occurs, with resultant necrosis in heavily involved tissues, and a mixed cell reaction develops, including polymorphs, macrophages, lymphocytes and plasma cells.

The rickettsial diseases include **endemic (murine) typhus**, caused by *R. mooseri* and transmitted by the rat flea; **epidemic typhus** (*R. prowazeki*) and **trench fever** (*R. quintana*) which are spread by the body louse; the **spotted fever group** (*R. rickettsi*, etc.) transmitted from various animals to man by the bites of infected ticks or mites, and finally **scrub typhus** (*R. tsutsugamuchi*), transmitted from rodents to man by a mite. Epidemic and endemic typhus are of world-wide distribution: the epidemic disease occurs in crowded louse-infested communities, and is common in times of war, earthquakes and other major disasters. Man is the only known reservoir of infection of *R. prowazeki*, which can persist for years as a latent infection and cause relapse ('*recrudescence* typhus' or *Brill-Zinsser disease*): such cases are responsible for fresh outbreaks.

Various forms of spotted fever are related to particular localities.

The rickettsial diseases vary in their severity and

pathological detail: in all, the small blood vessels are involved, and lesions tend to result especially in the brain, heart and skin. Infected material is particularly dangerous to laboratory workers, and diagnosis is usually made by demonstrating a rising titre of antibody, either in the patient or in laboratory animals inoculated with the patient's blood, etc. Only *R. quintana* has been cultured successfully in cell-free media.

Q fever is a typhus-like illness caused by the *Coxiella burnetii* which closely resembles the rickettsiae but differs from them in its antigenicity and in being much more resistant to drying, etc. and in being capable of both intracellular and extracellular growth. It is a parasite of domesticated animals of worldwide distribution and man is infected by inhalation of droplets while attending to animal births or by drinking infected milk, etc. Q fever usually presents as a 'non-bacterial' pneumonia, although lesions may occur in the brain and other organs. *Cox. burnetii* may also colonise the valves of the heart, producing a form of infective endocarditis.

Diagnosis is usually based on a rising titre of antibody, but demonstration of *Cox. burnetii* in the blood by guinea-pig inoculation is sometimes necessary.

Mycoplasmal infections

Mycoplasmas are very small filamentous or coccobacillary micro-organisms which lack a cell wall but can be grown in cell-free media and are classed as bacteria. They are distributed widely and are pathogenic to many animal and plant species. In man only one species, *Mycoplasma pneumoniae*, has been shown conclusively to be pathogenic, although other mycoplasmas have been isolated from the lesions of various other diseases. A major difficulty arises from their ubiquity and the consequent contamination of culture media; they can pass through bacteria-retaining filters and are also liable to contaminate cell cultures used in virology and for other purposes.

Mycoplasma pneumoniae is the cause of one form of 'non-bacterial' pneumonia, which is endemic in most parts of the world and also occurs as outbreaks, particularly in children. The organism dis-

seminates in the body and may cause a meningo-encephalitis. The immune response includes the production of an antibody which cross-reacts at low temperatures with a human red cell antigen, and is responsible in some cases for acute haemolysis.

Chlamydial infections

The chlamydiae are a group of spherical micro-organisms intermediate in size between the larger viruses and bacteria. They are obligatory intracellular parasites, but otherwise resemble bacteria far more closely than viruses. The vegetative form multiplies by binary fission, and infection is spread by a smaller compact spore-like form (elementary body) which can survive, but not divide, extracellularly.

These organisms are enzootic in certain birds, including the psittacines (parrot family), and also cause infections in sheep, goats and cattle. In man, they are responsible for the sexually transmitted disease **lymphogranuloma inguinale**, for eye infections, the most important being **trachoma**, and for pulmonary infection (**ornithosis**) which results from inhalation of the organism. The initial reaction to chlamydial infection is granulomatous, with accumulation of macrophages and lymphoid cells, necrosis, formation of granulation tissue and scarring. In lymphogranuloma inguinale a small ulcerating primary lesion develops in the genitalia, but the draining lymph nodes become grossly involved and prolonged suppuration and extensive scarring result. Similar lesions occur extragenitally in cat-scratch disease, but the causal agent is uncertain.

Both antibodies and cell-mediated immunity develop in chlamydial infections, the latter probably being the more important in the elimination of the infection. These diseases are considered more fully in the appropriate systematic chapters.

Yeasts and fungi (*Eumycetes*)

Yeasts and fungi are primitive eukaryotic micro-organisms which are now usually classified as neither plant nor animal. They are mainly saprophytic and make an important contribution to the breakdown of dead animal and plant tissues. Only a few of the very many known species are pathogenic to man, and with some exceptions the lesions are superficial and not serious. Good examples of such infection are athlete's foot and thrush (candidiasis). However, in drug addicts, severely ill patients, and particularly in those with T-cell deficiency or on immunosuppressive therapy, some of the fungi can cause more extensive or even systemic infections.

Fungi grow typically as filamentous branching hyphae which form an interlacing mycelium; they produce spores, commonly on projecting (aerial) hyphae. Yeasts consist of simple spherical or ovoid cells which multiply by budding, but the distinction from fungi is not sharp, for the so-called dimorphic fungi can assume the form of either hyphae or yeasts, depending on the environmental conditions.

In general, superficial infections with yeasts or fungi promote a mild inflammatory reaction. The reactions to deeper and more extensive infections vary considerably depending on the nature of the parasite and the host responses: they include necrosis, abscess formation, granulation tissue, and aggregation of macrophages, lymphocytes and plasma cells. In some instances, an epithelioid and giant-cell reaction results in appearances similar to those of tuberculosis, and various hypersensitivity reactions may contribute to the pathological changes. Host defence appears to be mediated largely by cell-mediated immunity.

In superficial infections, diagnosis can often be made from the appearance of the lesion, supported by microscopy of skin or mucosal scrapings, but cultivation on suitable media is sometimes necessary.

Some examples of yeast and fungal infections are described briefly below.

Candidiasis. *Candida albicans* is normally present in the mouth and intestine, and on the surface of moist areas of the skin. Superficial invasion and proliferation results in white patches (**thrush**) consisting of yeast forms and elongated cells (pseudohyphae), with mild inflammation of the affected tissue. It occurs in the vagina, particularly in pregnancy, and in the mouth, particularly in infants and in patients on oral antibiotic therapy. More extensive local and systemic infections occur in debilitated, immunodeficient or immunosuppressed patients. *Muco-cutaneous candidiasis*, affecting principally the face, scalp and mouth, is a chronic and extremely disfiguring condition when it affects children with various grades of T-cell deficiency. In some cases, administration of transfer factor (p. 131) has been followed by remarkable and sometimes prolonged remission. Oesophageal thrush (Fig. 19.18, p. 599) is a common finding at necropsy in subjects who have died following a chronic debilitating disease.

Systemic candidiasis is rare; the lesions consist of multiple small abscesses, resembling those of pyaemia, and are usually most numerous in the kidneys.

Aspergillosis. The spores of various species of

aspergillus, which are filamentous fungi (Fig. 16.32, p. 481), are present in the atmosphere, and large numbers are inhaled, particularly by agricultural workers: clinical infection is, however, uncommon. It is largely confined to the bronchi and lungs and is usually due to *Aspergillus fumigatus*. In most cases there are predisposing factors, such as steroid therapy or the presence in the lungs of bronchiectatic or old tuberculous cavities in which large aspergillus colonies may develop. The fungus may be more aggressive, and produce suppurating and granulating lesions in the lungs: occasionally it invades blood vessels, causing septic thrombosis, and a pyaemic condition.

The immune response includes antibodies and cell-mediated immunity, and complex hypersensitivity reactions may result. In Northern Sudan, a tumour-like granuloma occurs in the paranasal sinuses and orbit.

Histoplasmosis. This is caused by the dimorphic fungus, *Histoplasma capsulatum*, and occurs in many parts of the world, including some parts of North America and Europe. In Africa, most cases are due to *H. duboisii*. Infection usually results from inhalation of spores, which are present in soil and in the faeces of dogs, cats, rodents, bats and birds. In man, pulmonary lesions are most common; they may be single or multiple and usually heal and become calcified. The hilar lymph nodes are often involved. Progressive lung disease sometimes develops and resembles chronic pulmonary tuberculosis in its effects. The organism multiplies in macrophages, in which it is seen as multiple small yeast-like bodies with a double contour: aggregates of macrophages undergo caseous necrosis. In disseminated infection the macrophages throughout the body are colonised and large lesions occur in the liver, spleen, adrenals, marrow, etc.

In areas of high incidence, skin tests with an extract of *H. capsulatum* elicit a delayed hypersensitivity reaction in most individuals, indicating that a

high percentage of the population has developed immunity.

Cryptococcosis is caused by the yeast *Cryptococcus neoformans*, which grows in the droppings of pigeons and other birds. Infection in man occurs sporadically throughout the world. It probably results from inhalation of the organism, and produces a localised granulomatous pulmonary lesion: this may heal or extend, and spread may occur by the bloodstream, resulting in widespread granulomatous lesions mostly in the skin, lymph nodes and bones. A chronic meningitis also occurs, in which masses of yeast are seen macroscopically as gelatinous material.

Sporotrichosis is caused by a dimorphic fungus, *Sporothrix schenckii*, which is saprophytic on plants and is present in soil. Infection in man results from accidental inoculation of wounds or minor trauma. Chronic suppurating lesions develop locally and along the line of the draining lymphatics, but systemic infection is rare.

Protozoal and metazoal parasites

Most of the serious diseases caused by protozoan and metazoan parasites are now largely confined to tropical and sub-tropical countries, where they are responsible for an enormous amount of suffering. Because of the increase in world travel, however, these diseases are now encountered more often in visitors and immigrants to temperate areas, and an awareness of this is of major diagnostic importance.

The nature of the parasites, their life cycles and the features of the diseases they cause, are so varied that few useful generalisations can be made, and accordingly the more important individual diseases are described briefly in later chapters, under the systems in which they produce their major effects.

Further Reading

-
- Christie, A. B. (1974). *Infectious Diseases: Epidemiology and Clinical Practice*, 2nd edn., pp. 1095. Churchill-Livingstone, Edinburgh, London and New York. (A highly readable text, dealing with all aspects of infectious disease.)
- Collee, J. G. (1976). *Applied Medical Microbiology*, pp. 121. Blackwell Scientific, Oxford. (A short text for students.)
- Davis, B. D., Dulbecco, R., Eisen, H. N., Guinsberg, H. S. and Wood, W. B. (1973). *Principles of Microbiology and Immunology*, 2nd edn., pp. 1562. Harper and Rowe, New York. (A comprehensive, well-written text.)
- Duguid, J. P., Marmion, B. P. and Swain, R. H. A. (1978). *Medical Microbiology, Vol. 1*, 13th edn., pp. 666. Churchill-Livingstone, Edinburgh, London and New York. (A book for undergraduate and postgraduate students.)
- Freeman, Bob A. (Ed.) (1979). *Burrows' Textbook of Microbiology*, 21st edn., pp. 1138. Saunders,

- Philadelphia, London and Toronto. (A multi-author text with a major contribution by the editor. Readable and well-illustrated.)
- Timbury, M. C. (1978). *Notes on Medical Virology*, 6th edn., pp. 138. Churchill-Livingstone, Edinburgh, London and New York. (A short text for students.)
- Tyrrell, David A. J., Phillips, Ian, Goodwin, Stewart C. and Blowers, Robert (1979). *Microbial Disease: the use of the laboratory in diagnosis, therapy and control*, pp. 340. Edward Arnold, London. (A clearly-written book which relates clinical problems with laboratory practice.)
- Wilson, G. S. and Miles, A. A. (Eds.) (1975). *Topley and Wilson's Principles and Practice of Bacteriology, Virology and Immunity*, 6th edn., Vols. 1 and 2, pp. 2848. Edward Arnold, London. (A comprehensive text for practising bacteriologists).

Disturbances of Blood Flow and Body Fluids

Disturbances of the flow of blood are intimately associated with lesions which affect the functioning of the heart and blood vessels: such lesions will be considered systematically in later chapters. Meanwhile it is useful to outline the main features of disturbances in

total and local blood flow, the processes of thrombosis and clotting of the blood, and the disturbances in composition and volume of the body fluids. Accordingly, this chapter provides a general account of these phenomena.

Changes in Flow and Distribution of the Blood

Increase in total blood flow

This occurs when a sufficient number of arterioles relax to result in significant increase in the rate of passage of blood from the arterial to the venous compartment of the circulation. Physiological examples include the active hyperaemia in the skeletal muscles during physical activity and in the splanchnic circulation during digestion of a heavy meal. Pathological conditions causing an increase in total blood flow include the following.

(a) **Hypoxia**, which consists of significant fall in the amount of oxygen delivered to the tissues. This occurs in anaemia, i.e. a reduction in the amount of haemoglobin in the blood. In severe anaemia, cardiac output increases, but not enough to compensate for the reduced oxygen carrying capacity of the blood, and the tissues suffer from **anaemic hypoxia**.

Hypoxia occurs also when, as a result of various abnormalities of pulmonary function, the arterial blood is not fully oxygenated (**hypoxic hypoxia**). In lesions which interfere with pulmonary ventilation, the situation is complicated by increased P_{CO_2} of the blood, which, together with lowered P_{O_2} , is termed **asphyxia**. Congenital abnormalities of the heart or great vessels which result in mixing of venous and arterial blood can also cause hypoxic hypoxia.

Increased cardiac output is a feature of these various conditions, but it does not, of course, occur in heart failure, in which **ischaemic hypoxia** results from diminished perfusion of the tissues due to failing capacity of the heart to maintain the circulation.

(b) **Increased metabolic activity**. The general body metabolism is increased in hyperthyroidism (thyrotoxicosis), in fever, and convalescence from severe injury. The increased metabolism in these conditions is associated with an increased total blood flow.

(c) **Arterio-venous shunts**. A single large communication (fistula) between an artery and vein, such as sometimes results from trauma, allows the transfer of part of the cardiac output to the venous side of the circulation, and so reduces the amount of arterial blood available for tissue perfusion.

(d) **Extensive active hyperaemia**. In generalised inflammatory conditions of the skin, the active hyperaemia is sufficiently extensive to cause a significant increase in total blood flow. Similarly, a chronic increase in total blood flow occurs in Paget's disease affecting several large bones. In this condition, the marrow cavity of the affected bones is replaced by vascular granulation tissue, with local increase in blood flow. Also, there is persistent reflex active hyperaemia of the overlying skin and

soft tissues, and the total blood flow is consequently increased (see Singer *et al.*, 1978).

(e) **Liver failure.** The cause of increased blood flow in liver failure is uncertain: it may be due to the vasodilator effects of accumulated metabolites, or of compounds absorbed from the gut, which are normally removed from the blood by the liver cells.

In these various conditions, increased cardiac output is associated with a lowering of arteriolar tone: the pulse is bounding (of high amplitude) and the skin is warm and pink. The mechanism of these changes is complex and not fully understood: the autonomic nervous system, vasomotor centres, adrenal cortex and medulla, local effects of tissue metabolites, baro- and chemo-receptors, are all involved. If long continued, as in untreated hyperthyroidism, the increased work of the heart is likely to lead to '*high-output*' cardiac failure, particularly in older people and especially if the heart is already handicapped by coronary artery disease or other abnormalities.

Locally increased blood flow

The outstanding example of a pathological increase in local blood flow is **acute inflammation**, in which arteriolar dilatation results in active hyperaemia (p. 45) and the characteristic warmth and erythema of the inflamed tissue. Active hyperaemia occurs also following a **period of temporary obstruction of the circulation**: this is important when the local circulation is arrested to facilitate a surgical operation, e.g. on a limb, for hyperaemia develops gradually, and small vessels which do not bleed immediately after the circulation is restored may subsequently do so.

Reduction in total blood flow

This is a feature of **heart failure**, in which the heart is incapable of maintaining the normal output. The condition may occur acutely, usually as a result of myocardial infarction, or chronic heart failure may result from inadequate function of the myocardium, usually due to coronary artery disease or to increased workload as in valvular lesions or pulmonary or systemic arterial hypertension. Chronic heart failure is often progressive; the heart is incapable initially of supplying the increased

output required during physical activity, etc., but eventually it may fail to maintain an adequate circulation even at rest.

Reduced cardiac output is also the major feature of the acute condition of **shock**, in which grossly inadequate tissue perfusion can be fatal (pp. 260–5).

The cardiac output is also reduced in states of **general metabolic depression**, the commonest example being hypothyroidism, but in this instance it simply reflects the reduced requirements for tissue perfusion and is not of pathogenic importance.

The serious effects of heart failure are due very largely to **defective tissue perfusion**, which impairs the functions of all the organs. There are, however, two important structural effects: one is **general venous congestion**, from which the term **congestive heart failure** is derived: it is described below. The other is an increase in extravascular fluid, giving rise to **oedema**, which is described on pp. 253 *et seq.*

Local reduction in blood flow (local ischaemia)

This is of extreme importance since it accounts for a high proportion of cardiac and cerebral disease. Reduction of flow is usually due to **arterial narrowing**, or **complete obstruction by thrombosis or embolism**. These latter processes are described on pp. 235–45, and local ischaemia on pp. 245–51.

Local ischaemia can result also from **venous obstruction**, when it is accompanied by local venous congestion and commonly by oedema.

Venous congestion

When the heart fails to expel the normal amount of blood, arteriolar tone in general increases and a greater proportion of the blood accumulates in the venous compartment, which is readily distensible. This, together with an increase in blood volume (the mechanism of which is poorly understood) causes the veins to become engorged with blood. **Systemic venous congestion**, i.e. engorgement of the systemic veins, is most severe when the failure is predominantly of the right ventricle, as occurs in narrowing (stenosis) of the pulmonary valve orifice and in various diseases of the lungs which interfere with pulmonary blood flow.

Pulmonary venous congestion develops when there is failure of the left ventricle, as in many cases of coronary artery disease or systemic arterial hypertension: it occurs also when mitral valve stenosis restricts the flow of blood into the left ventricle, and may be present for many years without the development of heart failure. In both conditions, there is a rise in pulmonary arterial pressure due to hypertrophy and increased tone of the pulmonary arterioles, often leading to right ventricular failure and systemic venous congestion.

Venous congestion may also be localised to parts of the systemic circulation as a result of obstruction to the venous outflow. Such **localised venous congestion** is commonly seen as a result of thrombosis of the leg veins, often extending up to and involving the femoral vein. It occurs in the spleen and gastro-intestinal tract when portal venous flow is obstructed, as in cirrhosis of the liver. Various other veins may be obstructed, either by thrombosis or by pressure or constriction by a tumour or by scar tissue.

Systemic venous congestion

As explained above, this usually results from heart failure and, depending on the nature of the heart lesion, may be acute or chronic. In both instances, the outlook depends on the reversibility or otherwise of the cardiac failure: if this persists for long, the morphological changes of chronic venous congestion are striking, but it must be emphasised that *the congestive element is less important than the inadequate tissue perfusion of heart failure*.

The systemic veins can dilate to accommodate more blood without an immediate rise of venous pressure, but as the congestion increases, the pressure rises. This may be demonstrated directly by venous catheterisation, but commonly it is apparent from pulsation of the veins in the neck when the patient is sitting or standing. Normally, the neck veins in these postures are partly collapsed and do not pulsate visibly, the pressure in them being slightly below atmospheric. When venous pressure rises, however, the veins in the lower part of the neck are distended, and they pulsate at about the

level where the blood is at atmospheric pressure.

Because of the reduced blood flow in heart failure, the degree of oxygen dissociation in the capillaries is greater than normal, and in vascular tissues there may be sufficient reduced haemoglobin to give the purple-blue colour of **cyanosis**: this is seen, for example, in the lips and buccal mucosa. When there is also systemic venous congestion, the distension of the venules and capillaries with sluggishly-flowing oxygen-deficient blood increases the degree of cyanosis. Venous congestion may be present without oedema, but oedema usually accompanies severe congestive heart failure: it is most marked in the lower parts of the body, and chronic hypoxia and the increased venous pressure are probably both contributory factors.

Structural changes of systemic venous congestion. Apart from gravity-dependent oedema (p. 255), the structural changes in systemic venous congestion are most obvious in the abdominal viscera. The liver may be moderately enlarged and is often tender and palpable. Microscopically, the centrilobular veins* are dis-

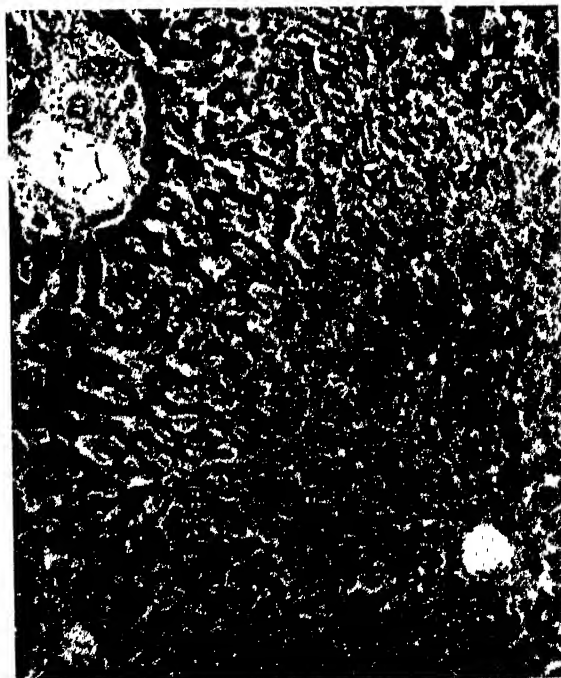


Fig. 9.1 Liver in chronic venous congestion, showing centrilobular atrophy and disappearance of liver cells accompanied by dilatation of sinusoids (rt. side of figure). $\times 105$.

* The relationship between the traditional liver *lobule* and the *acinus*, the newer concept of the structural unit of the liver, is described on p. 661.

tended and the central part of each lobule consists of distended sinusoids, the hepatocytes having undergone atrophy and disappeared (Fig. 9.1). Macroscopically, this results in accentuation of the lobular pattern, the dark, congested centrilobular areas contrasting with the paler, sometimes fatty peripheral lobular cells (Fig. 9.2). Because of its similarity to the

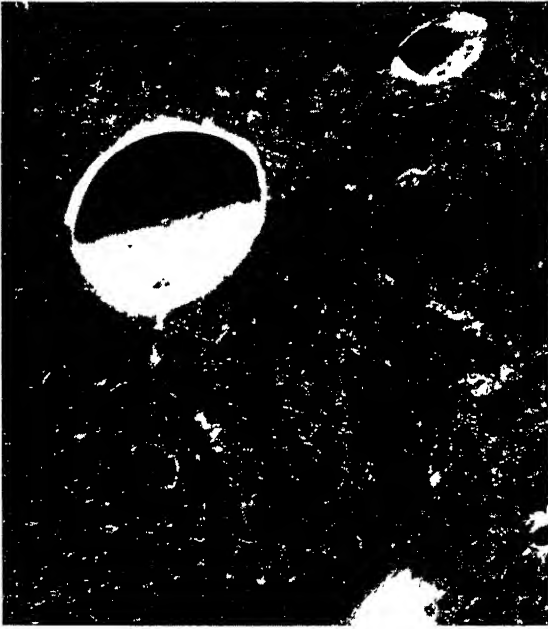


Fig. 9.2 The cut surface of the liver in chronic venous congestion. The congested centrilobular zones are dark, and contrast with the pale peripheral-lobular zones, giving the nutmeg-like appearance. $\times 1.8$.

surface of a nutmeg cut longitudinally, this appearance has long been described by pathologists as 'nutmeg liver'. In some cases, and particularly when there have been recurrent periods of congestive heart failure, centrilobular fibrosis occurs and nodules of hyperplastic parenchyma result from compensatory proliferation of surviving hepatocytes. The liver then appears diffusely irregular (Fig. 9.3): although commonly termed *cardiac cirrhosis*, these changes differ from true cirrhosis and do not progress to liver failure.

The **spleen** may be enlarged up to 250 g. It feels firm and maintains its firmness and shape on slicing, little blood escaping from the cut surface. The red pulp is congested and appears almost black: the Malpighian bodies may be visible as contrasting pale spots. Microscopy



Fig. 9.3 Nodules of hyperplasia in the liver in chronic venous congestion, giving the irregular appearances of so-called cardiac cirrhosis. $\times 1$.

shows congestion of the venous sinuses in the red pulp, with some thickening of the reticulin framework and atrophy of the medullary cords (Fig. 9.4). More marked congestion of the



Fig. 9.4 Chronic venous congestion of the spleen. The vascular sinuses are distended with blood, and the intervening medullary cords are relatively inconspicuous. $\times 250$.

spleen is seen in portal venous hypertension (p. 692).

The kidneys may be slightly enlarged and the medulla is particularly dark and congested; congestion is less obvious in the cortex, and appears as dark radial streaking (Fig. 9.5).



Fig. 9.5 Kidney in chronic venous congestion, showing the intense vascular engorgement, particularly of the medulla. $\times 0.7$.

These changes in the abdominal viscera are without serious effects: there may be mild or sub-clinical jaundice, and some red cells and protein in the urine, but the underlying condition of cardiac insufficiency is far more important.

In venous congestion of the lungs, the pulmonary venules and alveolar capillaries are engorged with blood (Fig. 9.6) and their walls become thickened. Red cells escape into the alveoli, sometimes resulting in bloodstained sputum, but many of them are broken down by alveolar macrophages, which come to contain large amounts of haemosiderin. The macrophages accumulate in the alveoli around respiratory bronchioles (Fig. 10.10, p. 279) and as haemosiderin is gradually released, the reticulin

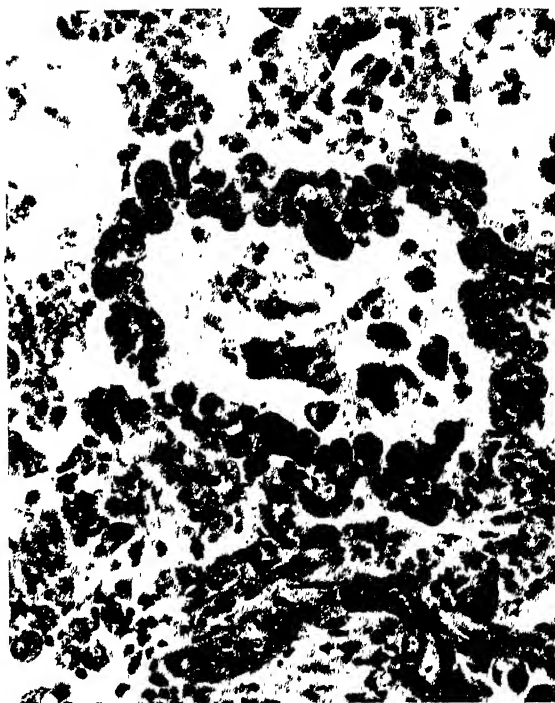


Fig. 9.6 Chronic pulmonary congestion, showing thickening of the alveolar walls, capillary congestion, and free (iron-containing) macrophages in the alveolar spaces. $\times 210$.

fibres in the alveolar walls become encrusted with it and fibrous thickening occurs. These changes result in increased firmness and give the lung a brown appearance (*brown induration*). The fine structural changes of pulmonary venous congestion are described on p. 457.

Many patients with chronic pulmonary congestion suffer from attacks of pulmonary oedema. They also develop pulmonary hypertension with its associated vascular changes (pp. 454-7). The iron-laden macrophages may be found in the sputum: they have been termed 'heart failure' cells, but are often present in pulmonary venous congestion, e.g. in mitral stenosis, for years before heart failure supervenes.

Local venous congestion

As mentioned above, this results from mechanical interference with the venous drainage of blood from an organ, limb, etc. The effects depend on the rapidity, degree and duration of obstruction and also on the local vascular arrangements.

Acute venous obstruction, e.g. by thrombosis or by a ligature does not usually cause com-

plete arrest of blood flow because in most parts of the body there is sufficient venous anastomosis to carry the blood away from the drainage area affected. In a few places, e.g. in the intestine, venous anastomosis is inadequate: the tissue becomes swollen, engorged with blood, and haemorrhagic due to rupture of small vessels. Ischaemic necrosis (venous infarction, p. 248) then develops.

In most sites, acute venous obstruction has less serious effects, and acute congestion either subsides or becomes chronic. This is illustrated by thrombosis of the deep veins of the leg, which is the commonest example of local venous obstruction, and often extends up to the femoral vein and even beyond. The limb may become cold, cyanosed and oedematous, but there is nearly always sufficient anastomosis to prevent infarction. The effects tend to subside gradually, partly because the anastomotic channels increase in calibre, and partly because the size of the thrombus is reduced by contraction and by digestion by plasmin. The lumen of the occluded vessel is often largely restored, and blood flow increases. Eventually, organisation and recanalisation of residual thrombus may further restore blood flow, but in spite of these changes, deep vein thrombosis, if extensive, sometimes results in chronic venous obstruction and persistent oedema of the limb. Venous valves may be put out of action if they are caught up in an organising thrombus, and this may also be a source of persistent trouble.

Chronic venous obstruction may result from thrombosis, as mentioned above, or from compression or invasion of a vein by tumour, or constriction by fibrous tissue. When obstruction develops gradually, collateral veins enlarge (Fig. 9.7) and drainage is often well maintained.



Fig. 9.7 Infra-red photograph showing enlargement of superficial veins to establish collateral circulation in a patient with obstruction of the inferior vena cava. (Dr. G. Watkinson.)

In chronic portal venous obstruction, which is an important effect of cirrhosis of the liver, the veins connecting the portal venous tributaries with systemic veins become enlarged and help to drain the portal system, but one such group of anastomotic veins, which run longitudinally in the submucosa of the lower oesophagus (Fig. 19.22, p. 604), is liable to rupture, causing serious or fatal haemorrhage.

Haemostasis and Thrombosis

It is essential that the blood should remain fluid within the cardiovascular system, and yet should be capable of *local haemostasis* by forming a solid adherent plug to prevent excessive bleeding from an injury to a vessel wall. The vital importance of these properties of blood is reflected in the complexity of the systems involved.

The repair of vascular injury

Injuries of vessel walls can be classified as major, when tissue is torn or cut and blood vessels are severed, and as minor 'wear-and-tear' defects which result from normal activities. **The minor defects** are presumably due to injury or

loss of individual endothelial cells with consequent exposure of collagen, elastin, etc. Such lesions are repaired almost instantaneously by adherence of platelets, followed by growth of endothelium over the adherent platelets to restore the integrity of the vessel wall. The importance of platelets is illustrated by the spontaneous haemorrhages which occur from the microvessels in severe thrombocytopenia (a reduction in the number of platelets). The role of fibrin deposition is illustrated by the haemorrhages into joints (haemarthroses) in severe haemophilia, a hereditary condition in which the coagulation or clotting process (formation of solid fibrin) is defective. Evidently the small vessels in the joints are normally exposed to degrees of injury which, although minor, cannot be repaired by platelets alone.

More severe injury results in partial or complete severance of larger vessels. Loss of blood is diminished temporarily by **vasoconstriction**: platelets stick to the collagen fibres of the torn edge of the vessel wall and, together with deposition of fibrin strands, gradually build up to form a mass—the **haemostatic plug**, which may close the gap in the vessel and prevent further bleeding. Vascular endothelium extends to cover the haemostatic plug which is then gradually removed by the process of organisation and proliferation of smooth muscle cells, the sub-endothelial gap thus being permanently repaired.

The factors involved in haemostasis

As explained above, vascular defects are repaired initially by formation of a haemostatic plug. This is an example of *thrombosis* (intravascular formation of solid material or *thrombus* from the constituents of the blood). In order to understand such beneficial haemostasis, and also disorders which upset the normal balance between the fluidity of the blood and its capacity to undergo thrombosis, it is necessary to consider in more detail the major factors involved: these are the *platelets*, *vascular endothelium*, the *clotting or coagulation process* which leads to deposition of fibrin, and the *plasmin system* which digests fibrin.

Platelet function

In normal blood, platelets circulate as single disc-like fragments of cytoplasm lined by a

plasma membrane. They do not adhere to normal endothelium but, as stated above, if endothelium is lost they adhere to the exposed collagen fibres. Adherent platelets undergo structural changes and discharge the contents of their storage granules (Fig. 9.8)—the *platelet release reaction*; these include adenosine diphosphate (ADP), 5-hydroxytryptamine, platelet factor 3, thromboxane A_2 , stable prosta-



Fig. 9.8 Electron-micrographs of platelets. Above, free platelets in suspension, showing the dense storage granules. $\times 8000$. Below, two glomerular capillaries plugged by aggregated platelets, most of which have discharged their granules. $\times 3000$.

glandins, and a factor which stimulates proliferation of smooth muscle cells. These platelet products have at least three important effects. Firstly, thromboxane A_2 , 5-HT and some stable prostaglandins, e.g. F_1 , promote local vascular contraction. Secondly, ADP and thromboxane A_2 both cause free platelets to alter shape, throw out pseudopodia, and adhere to one another and to those already adherent to the vessel wall: the release reaction continues and the result is the rapid development of the platelet mass. Thirdly, platelet factor 3 (and also factors released from injured tissue) initiates the clotting mechanism: this results in deposition of strands of fibrin which reinforce the platelet plug. Fibrin and thrombin (another product of the clotting system) both tend to promote further deposition of platelets.

Platelet function may be studied in various ways, e.g. by assessment of the percentage of platelets in a given sample which adhere to a standard column of glass beads through which the blood is passed, or the aggregation of platelets in plasma following the addition of aggregating substances such as ADP. These techniques have shown that there are a number of syndromes in which, although platelet counts are normal, platelet function is abnormal. Conditions with deficient platelet function include uraemia, the primary thrombocytopathies and hereditary haemorrhagic telangiectasia (p. 557): increased platelet adhesiveness and aggregation have been found in diseases associated with thrombo-embolic phenomena such as ischaemic heart disease, peripheral vascular disease and venous thrombosis. In the puerperium and following surgical operations, platelet adhesiveness is increased, the effect being maximal around the tenth post-operative day. This contributes to the post-operative thrombotic tendency.

Prostacyclin and thromboxane A_2 . In the past five years intense interest has developed in the influence of prostaglandins on platelet function and thrombosis. This has been stimulated by the discovery of prostacyclin (PGI_2) and thromboxane A_2 , both of which are highly unstable prostaglandins with half lives of about 2 minutes and 30 seconds respectively. They are derived from arachidonic acid, a polyunsaturated long-chain fatty acid which is produced from linoleic acid and is a constituent of the phospholipid of cell membranes. All cells also possess in their membrane prostaglandin (PG) synthetase

(cyclo-oxygenase), which converts arachidonic acid into unstable cyclic endoperoxides. As shown in Fig. 9.9, the endoperoxides, in turn, can be converted into prostacyclin, thromboxane A_2 , and stable prostaglandins (PGE_2 , etc.). Vascular endothelium is rich in prostacyclin synthetase and so produces mainly prostacyclin, whereas platelets convert endoperoxides mainly into thromboxane A_2 .

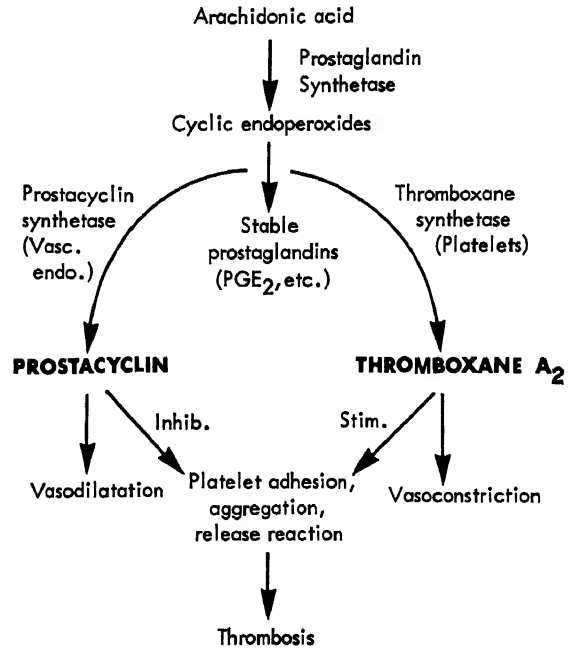


Fig. 9.9. The production of the unstable prostaglandins—prostacyclin and thromboxane A_2 —by vascular endothelium and platelets respectively, and their effects on the adhesion, aggregation and release reaction of platelets.

The role of thromboxane A_2 in promoting haemostasis is described above. There seems no doubt that it plays an important role in platelet aggregation and thrombosis. Inhibition of its production by aspirin in low dosage or imidazole derivatives (both of which inhibit thromboxane synthetase) renders platelets less readily adhesive and prolongs the bleeding time from a minor puncture wound. Prostacyclin has antagonistic effects: it causes vasodilatation and inhibits aggregation of platelets, disperses pre-formed platelet aggregates, and so inhibits thrombosis. It now seems very likely that these effects on platelets are mediated via the cyclic nucleotide system (p. 144), prostacyclin stimulating adenyl cyclase and thus increasing cAMP, thromboxane A_2 having the opposite effect. Platelet aggregation and the release reaction are promoted by a low cAMP and inhibited by a high cAMP level.

It has been proposed by Moncada and Vane (1979) that prostacyclin exerts both a local and sys-

temic control on platelet function. Its production by vascular endothelium may be of importance in preventing adhesion of platelets to normal endothelium. Mild physical or chemical injury to endothelium stimulates prostacyclin synthetase and so production of prostacyclin. Such injury may be sufficient to allow platelet adhesion to the damaged wall, thus effecting repair, but without the building up of a platelet aggregate, which is prevented by very low levels of prostacyclin.

The postulated systemic homoeostatic role of prostacyclin is attributed to its continuous release into the blood passing through the lungs. It is suggested that this maintains a level of cAMP in the free platelets which modulates their tendency to aggregate. There is, however, evidence that small loose aggregates of circulating platelets do sometimes occur, and this may be a cause of transient neurological symptoms in old people (p. 743).

These proposals are of potential importance in the use of drugs to influence thrombosis. For example aspirin, which is under trial in the prevention of coronary artery thrombosis, would be expected to inhibit thrombosis if given in low dosage, for this inhibits thromboxane synthetase; larger doses, which also inhibit prostaglandin synthetase, would be expected to interfere with production of both prostacyclin and thromboxane A_2 (Fig. 9.9), and thus have less or no inhibitory effect on platelet aggregation and thrombosis. There is some evidence in support of this (Masotti *et al.*, 1979).

The unstable prostaglandins are obviously of interest in relation to the thrombogenic theory of atheroma (p. 367).

The coagulation (clotting) mechanism

By clotting is meant the conversion of fibrinogen to solid fibrin. This is the result of a complex series of reactions involving sequential activation of a large number of clotting factors most of which have been purified, although the relative importance of the various parts of the system is not yet clear. Not only is the system complex, but most of the major factors have been numbered (I to XIII, Table 9.1) in the order of their discovery, and not in the order of their participation. The clotting process up to the activation of factor X can occur by two main routes, the *intrinsic* and *extrinsic pathways*. After this stage, there is a *common pathway* leading to the formation of fibrin. The complexity of the process is indicated by Fig. 9.10 which is a simplified scheme.

The **intrinsic pathway** results in clotting without the participation of factors released from

Table 9.1 International classification of the plasma coagulation factors (Roman numerals), together with their commonly-used names. The term 'factor VI', formerly applied to an intermediate product, is no longer used.

Factor I	Fibrinogen
Factor II	Prothrombin
Factor III	Tissue factor
Factor IV	Calcium
Factor V	Proaccelerin
Factor VII	Proconvertin
Factor VIII	Antihæmophilic globulin
Factor IX	Plasma thromboplastin component or Christmas factor
Factor X	Stuart-Prower factor
Factor XI	Plasma thromboplastin antecedent
Factor XII	Hageman factor
Factor XIII	Fibrin stabilising factor (plasma transglutaminase)

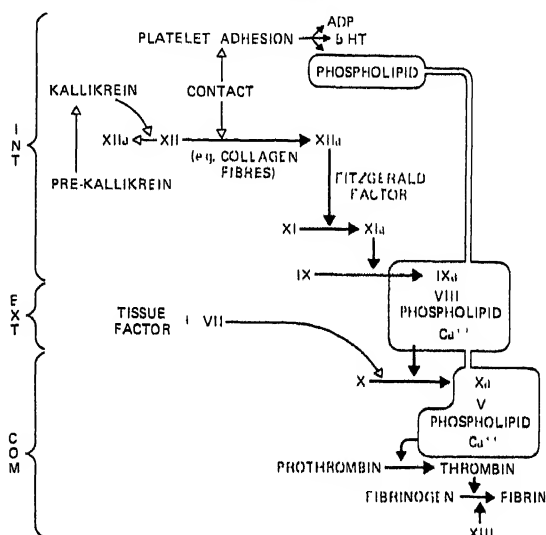


Fig. 9.10 The reactions involved in the clotting mechanisms. INT, intrinsic pathway; EXT, extrinsic pathway; COM, common pathway.

injured tissues. It occurs when blood is placed in a tube and is probably involved in minor vascular injury and when blood stagnates in a blood vessel. It is initiated by contact of factor XII (Hageman factor) with a foreign surface or with collagen. Activated factor XII (termed XIIa)* is an esterolytic enzyme which, together with another factor (Fitzgerald factor), converts factor XI to XIa; this, in turn, activates factor IX. Factor X is then activated by a reaction involving IXa, Ca^{++} , phospholipid and factor VIII (antihæmophilic factor).

The **extrinsic pathway** is activated by tissue injury. It is triggered by a lipoprotein complex

* The activated factors are indicated by the suffix 'a'.

(tissue factor 3 or thrombokinase), which activates factor VII: this in turn activates factor X.

The common pathway. Factor Xa, produced by either of the above routes, is a serine esterase which forms a complex with factor V and phospholipid in the presence of Ca^{++} : this complex converts factor II (prothrombin) to IIa (thrombin) which converts fibrinogen to fibrin monomer. Factor XIII then polymerises fibrin monomer to form insoluble fibrin.

The clotting process is influenced by a number of *amplifying and inhibiting reactions*. For example, in the intrinsic system activated Hageman factor (XIIa) converts prekallikrein (a component of the kinin system) to kallikrein, which activates more Hageman factor, while factor VII, which participates in the extrinsic pathway, is activated by a number of other factors—thrombin, XIIa, kallikrein, IXa and plasmin (see below). For the various activation steps there are specific inhibitors which modulate the clotting process and help to prevent inadvertent thrombosis.

The fibrinolytic (plasmin) system

This is shown diagrammatically in Fig. 9.11. Its activation results in the production of *plasmin*, a proteolytic enzyme which digests fibrin to soluble fibrin degradation products (FDP). Plasminogen is a β -globulin in the plasma. It is activated by a factor in vascular endothelium, by tissue factors, factor XIIa, and various bacterial products and chemicals. Plasmin digests not only fibrin, but also factors V, VIII and II (prothrombin): its activation and enzymic activity are controlled by a number of inhibitors in plasma, notably α_2 -macroglobulin and α_1 -antitrypsin.

Plasminogen binds selectively to polymerising fibrin and is then converted to active plasmin by activators which diffuse into the clot or thrombus where the effect of inhibitors is weak. Plasmin activity in the plasma is prevented by the presence of factors which both inhibit its formation and suppress its activity. It is, however, possible to prevent such inhibition by administration of urokinase or streptokinase. Increased plasmin activity of the plasma is found after exercise or emotional stress, and also following surgical operations and other trauma.

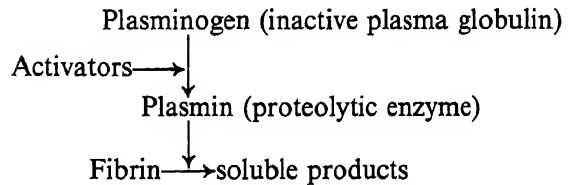


Fig. 9.11 The fibrinolytic enzyme system.

The fibrinolytic enzyme system is probably in dynamic equilibrium with the blood clotting system, the two acting together to maintain an intact and patent vascular tree. According to this hypothesis the coagulation and fibrinolytic systems may both be continuously active, the former laying down fibrin where needed on the endothelium to seal any deficiencies which may occur, and the latter removing such deposits after they have served their haemostatic function.

As explained in earlier chapters (e.g. p. 54) there are complex inter-relationships between the clotting, kinin, plasmin and complement systems, none of which can now be regarded in isolation.

Pathological Thrombosis

Thrombosis is defined as the formation of a solid or semi-solid mass from the constituents of the blood within the vascular system during life. Coagulation, i.e. deposition of fibrin, is involved in the formation of all thrombi except perhaps the minute deposits of platelets which maintain vascular integrity (p. 232). As already noted, the composition of thrombus is determined very largely by the rate of flow of the blood from which it forms.

Appearances and composition of thrombi

As a general rule, thrombus forming in rapidly flowing blood, e.g. in an artery, consists mainly of aggregated platelets, with some fibrin; it enlarges slowly and is firm and pale, varying from greyish white to pale red, and is commonly called *pale thrombus*. The proportion of fibrin deposited in pale thrombus depends partly on the rate of blood flow, to which it has an in-

verse relationship. At the other extreme, thrombus forming in stagnant blood, e.g. adjacent to a complete occlusion of a blood vessel, is indistinguishable from blood which has been allowed to clot *in vitro*: the thrombus is soft, dark *red*, gelatinous and consists of strands of fibrin lying among the elements of whole blood (Fig. 9.12). It may retract from the vessel wall, revealing a smooth, shiny surface. Between these two extremes we have *mixed thrombi* which form in slowly flowing blood, usually in veins, and consist of alternating layers of platelet aggregates and red thrombus. The mixture may be intimate and only recognisable on microscopy: Fig. 9.13 shows such a thrombus in which spaces between masses of aggregated platelets are filled by a fibrin network containing leukocytes and some red cells. In other instances, veins may be filled with columns of red thrombus but with platelet aggregates at points of anastomosis. The formation of such thrombi is explained on p. 240. Except in recently formed thrombi, it is not easy to recognise ag-



Fig. 9.12 Red thrombus, consisting of strands of fibrin lying among red cells, leucocytes and platelets. In this instance the proportion of entrapped red cells is much lower than in blood clot, indicating that there has been some flow of blood during thrombosis. $\times 305$.

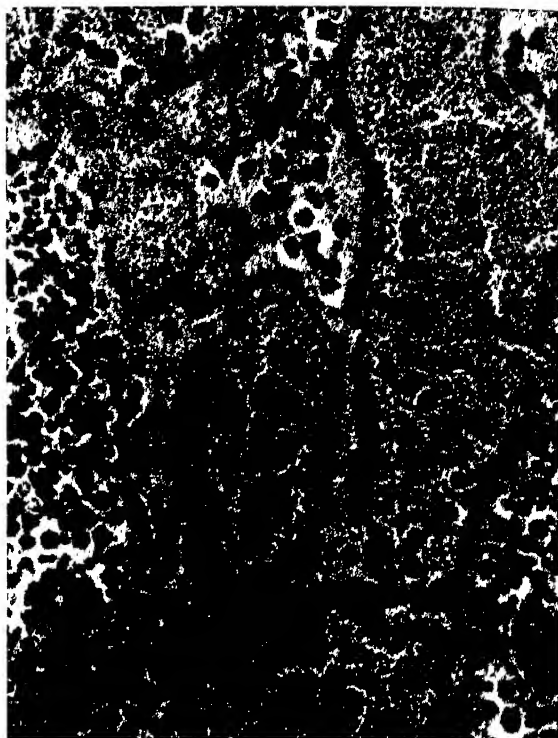


Fig. 9.13 Mixed thrombus, about 12 hours old, showing dense masses of granular material, composed mainly of platelets, fibrin strands and collections of leukocytes. $\times 336$.

gregated platelets, for they soon lose their outlines, presenting microscopically a granular or structureless appearance. Immunofluorescence studies and electron microscopy (Fig. 9.8) have, however, helped in their recognition.

Sites and predisposing factors of thrombosis

The three major predisposing factors are:

- (a) *local abnormalities in the walls of blood vessels or of the heart*
- (b) *slowing and other disturbances of blood flow*
- (c) *changes in the blood favouring platelet aggregation and fibrin formation.*

The roles of these three factors in the formation of thrombi in the heart, arteries and veins are considered below.

(a) Cardiac thrombosis

Thrombi may form on the walls of any of the chambers of the heart, and also on the valve cusps.

In the atria, thrombosis is commonest in the appendices, especially that of the right atrium, in cases of heart failure with atrial dilatation (Fig. 15.25, p. 417). Stagnation of blood is the important causal factor and this is accentuated by atrial fibrillation, which commonly occurs in such patients and is very liable to be complicated by atrial thrombosis: the thrombus is usually red and is moulded to the irregular wall of the atrium.

Rarely, small flattened globular thrombi form in either the atria or ventricles. They are pale, composed mainly of platelets and may show central softening. In mitral stenosis, a rounded thrombus may develop in the left atrium. It may exceed 3 cm in diameter and become detached to lie free: it is a rare cause of sudden obstruction of the circulation—the so-called ‘ball valve thrombus’. The vegetations which form **on the heart valves** in certain diseases are essentially thrombi. In rheumatic fever the valve cusps are damaged along the line of apposition, and deposition of platelets and fibrin results in the formation of minute pinkish-grey bead-like vegetations (Fig. 9.14 and Fig. 15.22, p. 413); in infective endocarditis the cusps are damaged by microbial infection and much more

fibrin is deposited, containing interspersed leukocytes; the vegetations are consequently larger, softer and more friable (Fig. 9.15). **In the ventricles**, mural thrombosis commonly occurs on the endocardium overlying an infarct (i.e. a patch of ischaemic necrosis of the heart wall—Fig. 9.16). Depending on the size of the infarct, the thrombus may be large or small. It forms a flat reddish—or, if older, a brown—patch attached to the endocardium. Probably the important factors in its formation are the disturbances in blood flow caused by lack of pulsation in the dead muscle and also diffusion of factor III (tissue thromboplastin) from the dead tissue.



Fig. 9.14 Rheumatic vegetation on a cusp of the mitral valve. The vegetation consists mainly of dense hyaline material (A) composed of fused platelets, and fibrin coagulum (B). $\times 60$.



Fig. 9.15 Bacterial endocarditis. The root of the aorta has been cut open and the wall folded back to display the large irregular thrombotic vegetations which have formed on, and are obscuring, the aortic valve cusps.

At necropsy, **clots formed after death** are usually to be found in the chambers of the heart. They are soft and dark red with a glistening surface and are not firmly adherent to the endocardium. Occasionally the red cells settle before coagulation occurs and the upper (usually anterior) part of the clot is then yellow and gelatinous. Thrombi may form rapidly as the circulation is failing immediately before death. They are yellow or pinkish with a glistening surface and have a somewhat stringy appearance (Fig. 9.17). Such **agonal thrombi** originate at



Fig. 9.16 Mural thrombus (*above*) which has formed on the ventricular endocardium over a myocardial infarct. Note the necrotic myocardium (*below*). $\times 120$.



Fig. 9.17 Agonal thrombus in the right ventricle, extending along the pulmonary artery. $\times \frac{1}{2}$.

the apex of the ventricle to which they are attached and may extend through the valve orifice. They are composed mainly of fibrin, which separates out from the sluggishly moving blood before death, and may occur in either or both ventricles, although they are commoner in the right side of the heart.

(b) Arterial thrombosis

Probably because of the rapid flow of blood, arterial thrombosis is uncommon in the absence of a local lesion of the vessel wall. In affluent communities, **atheroma** is by far the commonest predisposing local lesion. It consists of multiple patches of fibrous thickening and lipid deposition in the intima of arteries of various sizes. *In the aorta*, atheroma is commonly severe and results in gross distortion and unevenness of the wall. When blood is flowing smoothly in a normal vessel, the particulate elements are separated from the vascular endothelium by a layer of almost pure plasma, but atheromatous plaques, by causing irregularities of the wall, result in turbulent flow, and platelets can then impinge on the wall. This alone probably predisposes to thrombosis, but because of the rapid flow of blood, thrombosis is often not superadded. Frequently, however, atheromatous patches ulcerate, and thrombosis supervenes. In the aorta the thrombi are usually *mural*, i.e. the ulcerated atheromatous patch becomes coated by thrombus which does not extend to occlude the lumen (Fig. 9.18). Thrombosis also complicates atheroma in *medium-sized and smaller arteries*, particularly those supplying the heart and brain. Because of their relatively small calibre, which is further reduced by the atheromatous plaques, thrombi readily occlude these vessels completely and ischaemic necrosis commonly occurs in the deprived tissues. This is described later in the section on infarction and also in the appropriate systematic chapters. When there is gross **localised dilatation (aneurysm)** of the wall of the heart, aorta or other arteries, stagnation and eddying of the blood usually result in some thrombosis. The thrombus may have a laminated appearance and may come to fill the aneurysmal sac (Fig. 9.19). **Inflammatory lesions** in the walls of arteries (pp. 377–83) also cause thrombosis: contributory factors may be irregularities of the wall, injury to or loss of the vascular endothelium, and release of tissue thromboplastin. In **severe arterial hypertension**,



Fig. 9.18 Part of the abdominal aorta opened up to show a large thrombus which has formed over atheromatous patches. The dull, pale, shaggy thrombus consists mostly of fibrin and platelets.



Fig. 9.19 A large aneurysm of the aortic arch (c.f. size of heart). The aneurysmal sac has become largely filled by laminated thrombus.

necrosis of the walls of small arteries and arterioles is commonly followed by thrombosis.

(c) Venous thrombosis

Apart from varicosity of the leg veins, diseases of the veins are uncommon, and although venous blood flow is slow, occlusion of veins in general occurs less frequently, and is usually less serious, than occlusion of arteries. The most important exception is *thrombosis of the veins of the lower limbs*, which is very common in bedridden patients, and is the usual cause of serious or fatal pulmonary embolism (see below): less often, thrombosis occurs in the *pelvic veins*, and this also may cause pulmonary embolism.

Thrombosis of leg veins usually starts in deep veins of the leg, most often within the calf muscles (Fig. 9.20), from where it may extend progressively to the posterior tibial and popliteal veins, the femoral (Fig. 9.21) and iliac veins and occasionally to the inferior vena cava. In some instances, it may start more proximally than the calf, or several thrombi may form in the calf and thigh veins.

Extension of the thrombus is sometimes very rapid, the whole length appearing as soft red thrombus: this probably occurs when flow of

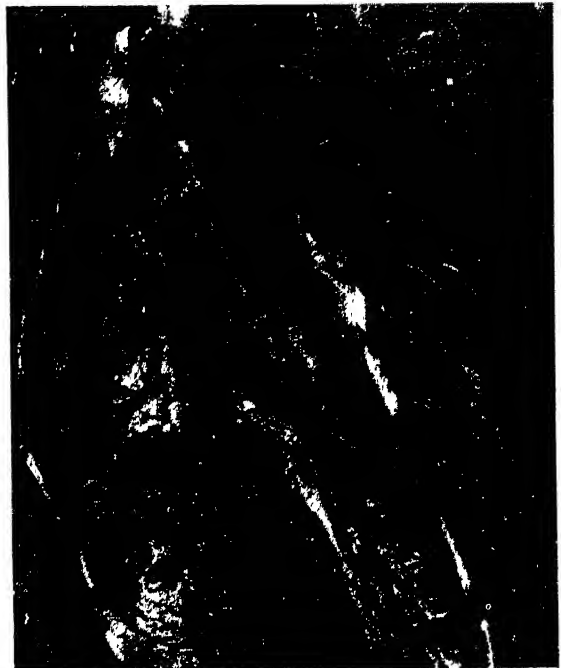


Fig. 9.20 Recently-formed red thrombus in the deep veins of the leg.

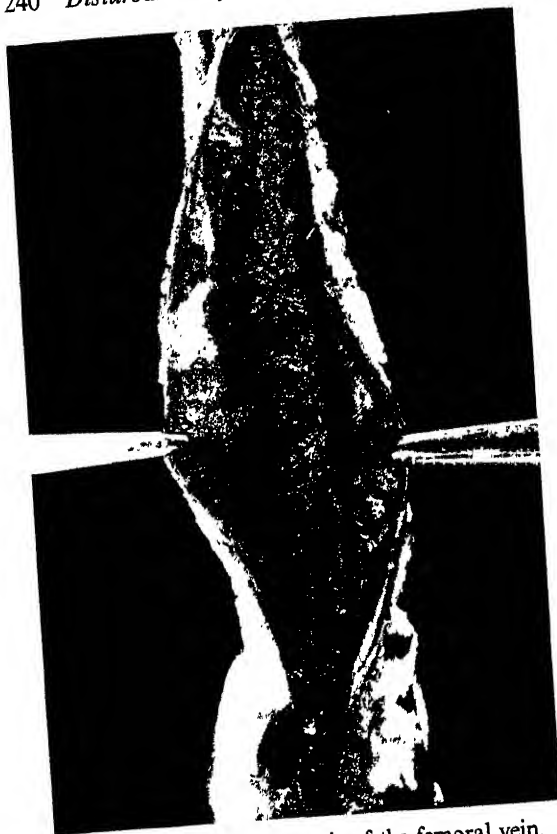


Fig. 9.21 Recent thrombosis of the femoral vein.

blood is very slow, and the thrombus resembles blood clot, being soft and red and readily detached from the vessel wall to reveal a red glistening surface or a paler, dull appearance due to deposition of platelets and fibrin. In other instances, probably where the blood flow is less sluggish, thrombosis extends more slowly, as depicted in Fig. 9.22. The initial thrombus formed in the calf veins occludes the lumen and for some distance proximal to the occlusion flow is virtually arrested. This column of blood is rapidly converted into red thrombus as far as the next proximal venous tributary. Blood from this tributary continues to flow into the affected vessel and for a time arrests the formation of red thrombus. However, platelets in the moving column of blood coming from the tributary are deposited on the proximal end of the red thrombus, which thus becomes capped with more slowly formed pale thrombus consisting mainly of platelets and strands of fibrin. This may eventually occlude the entrance of the tributary, again producing stagnation, and so red thrombus forms and extends proximally to the next tributary. So the process continues with

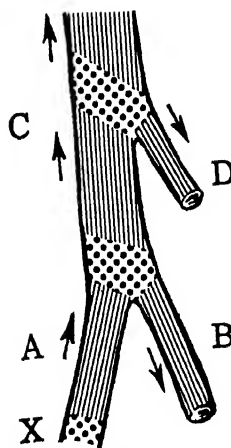


Fig. 9.22 Diagram showing the mode of extension of venous thrombosis. Thrombus occludes a small vein (A) at point X, and red thrombus (lined areas) rapidly extends in the stagnant column of blood up to the entrance of the next tributary (B), where platelet deposition forms a cap of pale thrombus (dotted areas): when this occludes the junction of A and B, red thrombus extends rapidly up to the entrance of the next tributary (C), and so on. Red thrombus also forms in each tributary as its entrance to the major channel is occluded. (Arrows show direction of spread of thrombosis.)

thrombus extending into larger, more proximal veins. Once a tributary has been occluded, red thrombus also forms in the stagnant blood within it and so the thrombus in the main venous trunks comes to have branches extending into the tributaries. Leg vein thrombosis tends to occur especially in patients lying immobile in the supine position, and impairment of blood flow by pressure on the calves appears to be an important predisposing factor. It is particularly common after an abdominal operation, severe injury, myocardial infarction, and in patients with congestive heart failure. There is an increased risk during pregnancy and following childbirth. Venous return from the lower part of the body is normally aided by muscular movements of the legs and by the pumping action which ensues from use of the abdominal muscles and diaphragm in respiration, and in all the above conditions immobility of the legs interferes with the normal flow. To avoid the pain of abdominal movements, patients who have had an abdominal operation tend to use mainly the thoracic muscles for respiration, and this is a further factor in impairing venous return from the legs. There is good evidence that thrombosis usually starts in the

small leg veins *during* surgical operations. Extension into the large veins usually occurs over the next two weeks, when platelet numbers and adhesiveness and prothrombin levels are highest, and as in childbirth and myocardial infarction, pulmonary embolism is commonest around ten days after the event. The main factors predisposing to venous thrombosis in congestive heart failure are venous stagnation and immobility.

Leg vein thrombosis and embolism (see below) of large or small pulmonary arteries is an exceedingly common finding at necropsy on middle-aged and old patients (in our experience over 30 per cent). Recent reports on the prophylactic use of repeated small doses of heparin indicate that this reduces considerably both venous thrombosis and pulmonary embolism, but in surgical cases such therapy is more effective if started at the time of operation. It is partly to prevent venous thrombosis that patients are encouraged to leave their beds as soon as practicable after operation, childbirth, etc., and while bedridden to carry out muscular exercises and to practise abdominal respiration.

Thrombosis of the pelvic veins after operation, etc., is less common than leg-vein thrombosis. It is seen especially after childbirth when the uterine blood flow diminishes considerably, predisposing to thrombosis in the hypertrophied uterine veins. Puerperal sepsis is also a predisposing factor in some cases. Pelvic venous thrombosis may also originate in haemorrhoids or in the prostatic venous plexus and is a complication of operations on the pelvic organs, particularly if there is sepsis. Extension to large veins, including the internal and common iliacs, may complicate pelvic venous thrombosis and fatal pulmonary embolism may follow.

Other causes of venous thrombosis include *malnutrition, severe debilitating infections and wasting diseases* such as cancer. When associated with these conditions it is sometimes called **marantic thrombosis** and in severely debilitated infants and young children may affect the superior longitudinal sinus (Fig. 9.32, p. 249). Venous thrombosis is also prone to occur in **some disorders of the blood**, for example leukaemia and polycythaemia vera (excessive numbers of red cells, leukocytes and platelets). **Inflammation of veins (phlebitis)** also promotes

thrombosis: this may occur as a *migrating thrombophlebitis*, affecting various veins throughout the body: it is usually of obscure aetiology (p. 391), but is sometimes associated with cancer of the internal organs, particularly the pancreas. The thrombi are usually firmly adherent and embolism is unusual. In *septic venous thrombosis*, however, fragments of infected thrombus may break away and give rise to pyaemia (p. 201).

(d) Capillary thrombosis

Thrombosis in capillaries and venules commonly occurs in severe acute inflammatory lesions. It is due partly to endothelial damage and partly to haemoconcentration, the thrombi being composed mainly of packed red cells.

In the Arthus reaction, in which thrombosis of small vessels is often prominent, the endothelial injury is attributable mainly to release of enzymes by neutrophil polymorphs (p. 153).

Fibrin thrombi can be found in the capillaries in some patients dying of disseminated intravascular coagulation (p. 264), although in some cases they are absent, presumably as a result of fibrinolytic activity.

The fate of thrombi

Like other abnormal digestible material deposited in the body, thrombus is removed by enzymic action, and the success of such removal depends on the degree of restoration of the lumen of the thrombosed vessel.

The processes involved in removal of thrombus are (a) contraction of the thrombus, (b) digestion by plasmin and the proteolytic enzymes of neutrophil polymorphs trapped in the thrombus, and (c) organisation, involving digestion by macrophages and formation of fibrovascular tissue. The relative importance and effects of these processes depend on the type of thrombus and the site of its formation.

Occlusive venous thrombi are usually formed mainly of soft red thrombus which contracts well. Where it remains attached to the vessel wall, it becomes invaded by fibroblasts and macrophages, along with capillaries which are probably derived from the plexus of vessels external to the internal elastic lamina (vasa vasorum).

This ingrowth of granulation tissue does not often occur around the whole circumference of the vessel, usually being confined to sites where thrombosis has caused secondary damage to the intimal endothelium. Elsewhere around the wall, fluid-filled pockets may form where thrombus has retracted from the vessel wall or has been digested by the local action of plasmin. This is particularly prominent in the thrombus around the valve cusps of the vein. The cells of the intimal endothelium migrate and proliferate rapidly to cover the free surface of the thrombus, and also penetrate into it. This results in both fragmentation of the thrombus into tiny endothelial covered nodules and in the formation of small capillary channels, many of which are probably blind-ending, while a few link up with capillaries growing into the thrombus from the vein wall. The thrombus is also partially resorbed by the action of macrophages and sometimes the centre is softened by the enzymes from groups of dead polymorphs which have migrated from the thin-walled vessels. In this way, by the joining up of pockets, by fragmentation, resorption and softening of thrombus, a lumen may be restored leaving a thickened fibrovascular intimal plaque or a meshwork of fibrous strands marking the site of granulation tissue ingrowth with subsequent fibrosis. Occasionally, however, and perhaps when the thrombus is especially dense and slowly formed, it remains adherent to the whole circumference of the vein: pockets are not formed and significant recanalisation fails to occur. The thrombus is then replaced by granulation tissue which becomes increasingly collagenous, the vein eventually being reduced to a solid, shrunken cord without a lumen.

Occlusive arterial thrombi are usually formed more slowly and are more dense than venous thrombi; they contain a higher proportion of platelets and fibrin and are less readily digested. Also arterial endothelium is a relatively poor source of plasminogen activator and, possibly because of these factors, formation of pockets between the thrombus and arterial wall occurs much less than in veins. Consequently, the thrombus remains largely in contact with the artery wall and its removal takes place mainly by organisation. Macrophages migrate into the margin of the thrombus and gradually digest it. This is accompanied by invasion by fibroblasts from the intima and by new capil-

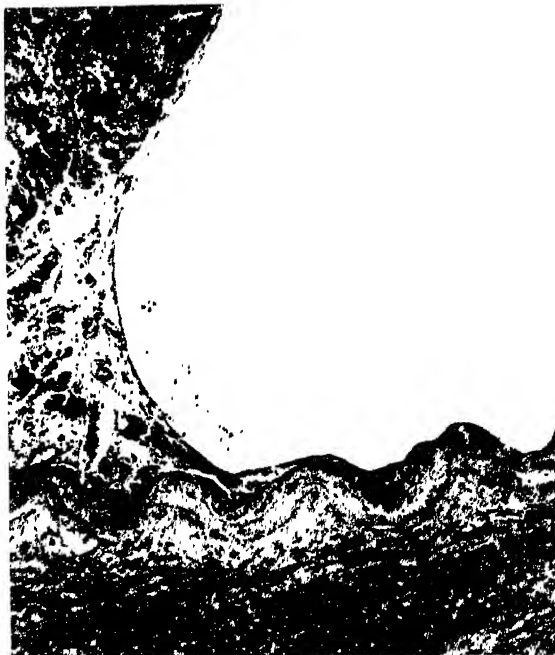


Fig. 9.23 Endothelium has grown over the surface of this partially organised thrombus adherent to an artery wall. $\times 40$.



Fig. 9.24 Part of an artery showing the results of organisation of thrombus. The lumen was originally to the left of the internal elastic lamina, which runs vertically near the right margin. The lumen is now filled with vascular fibrous tissue in which some of the new capillaries have enlarged and acquired muscle to become arterioles. $\times 115$.

laries which develop from the arterial endothelium, including that which grows over the ends of the thrombus (Fig. 9.23). The vasa vasorum probably play no part if the internal lamina is intact. The thrombus is thus gradually replaced by fibrovascular tissue and the length of vessel affected may eventually become a fibrous cord. The newly-formed capillaries anastomose and may develop into larger vessels extending through the length of the occlusion (Fig. 9.24), but such **recanalisation** does not often restore

effective blood flow.

Mural thrombus. Thrombus may form on a part of the wall of a vessel (or of the chambers of the heart) without extending to fill the lumen. Such mural thrombi are rapidly covered with endothelium from the surrounding intima. Fibrinolysis, fragmentation, resorption and enzymatic breakdown probably all play a part in removing some of the thrombus, while granulation tissue grows in from the underlying wall and organises the remainder.

Embolism

By embolism is meant the transference of abnormal material by the bloodstream and its impaction in a vessel. The impacted material is called an **embolus**. In most cases it is a fragment of thrombus, although fragments of material from ulcerating atheromatous plaques of the aorta quite commonly form emboli in more distal arteries. A fragment of a tumour growing into a vein may also break off and form an embolus, and there may be embolism of the capillaries by fat globules, air bubbles or even groups of parenchymal cells. The site of embolism will, of course, depend on the source of the embolus. Thus embolism of the pulmonary arteries and their branches is secondary to thrombosis in the systemic veins or in the right side of the heart. Rarely, where there is a patent foramen ovale, an embolus may pass from the right side of the heart to the left atrium and thus be carried to the systemic circulation; (*crossed* or *paradoxical embolism*). With this rare exception, emboli occurring in the systemic circulation are derived from thrombi formed in the left side of the heart, from thrombotic vegetations on the aortic and mitral valves, and from thrombi or detached portions of atheromatous plaques in the aorta or large arteries. Emboli carried from tributaries of the portal vein lodge, of course, in the portal branches in the liver.

Effects of embolism

Systemic arterial emboli. The results are simply those of mechanical plugging and vary

according to the site of the embolus, as described in pp. 245 *et seq.*

Pulmonary embolism is a very common and important event: it results from the detachment of a thrombus in a systemic vein, usually in the lower limb. Such thrombi form in conditions which have already been described (p. 240) and in any of them pulmonary embolism may result. It is most common around the tenth day after operation, and may cause sudden death. A large thrombus may become detached *en masse* and be carried to the right side of the heart, causing a sudden blockage of the pulmonary trunk or one of its divisions, death usually occurring at once or after a short period of pulmonary distress. Such fatal emboli are most often derived from the femoral and iliac venous trunk, characteristically forming a cylinder about 1 cm in diameter and as much as 30 cm long, which is found at necropsy coiled up like a snake in the pulmonary artery and right ventricle (Fig. 9.25). Depending on their size, less gross fragments of thrombus impact in the major or minor pulmonary arteries. When the patient has lived some time after the embolism, a varying amount of haemorrhagic infarction may be present in the parts supplied by the blocked vessels. Infarction, however, is never co-extensive with the area of distribution, and usually there is none.

Multiple small pulmonary emboli, impacting over a period of time, can rarely cause chronic pulmonary hypertension.

Septic emboli. With the widespread use of antibiotics, septic emboli, containing pyogenic bacteria, have become relatively uncommon.



Fig. 9.25 Massive pulmonary embolism. Thrombus from the femoral vein has become detached and impacted in the pulmonary trunk and its right and left branches, causing sudden death.

Where they impact, such emboli may cause abscess formation, and the multiple abscesses of pyaemia develop in this way. An infective embolus occasionally weakens the arterial wall and gives rise to an aneurysm—*mycotic aneurysm*. In various septicaemic and pyaemic conditions, capillaries here and there may be plugged by organisms, most frequently pyogenic cocci, or by impaction of a small fragment of infected thrombus, the organisms then growing along the capillaries. The number of bacteria seen in necropsy material may have been greatly increased by growth after death.

Embolism from tumours. This is of two kinds. One or a few cells of the tumour may enter the bloodstream and impact in a capillary in some distant organ. In other instances there may be growth of a tumour into a large vein, and a larger fragment may become detached and impact in a vessel, e.g. a branch of the pulmonary artery or of the portal vein. Both of these processes can result in metastatic tumours.

Fat embolism. Entrance of fat into the circulation results from laceration of veins surrounded by adipose tissue. It probably occurs

after all fractures with laceration of adipose tissue, in caisson disease (p. 779) and as a complication of a fatty liver. In most instances the phenomenon is of no clinical importance but when the amount of fat entering the circulation is large, as in fractures of long bones, the **fat embolus syndrome** may develop within the following 3 days. The syndrome includes mental confusion, fever, dyspnoea, tachycardia, a petechial rash and sometimes cyanosis, haemoptysis, coma and death. It appears to be due largely to hypoxia resulting from pulmonary fat emboli complicated by oedema and haemorrhage. Fat may, however, pass through the lungs into the systemic circulation and cause emboli in the brain, giving rise to multiple small haemorrhages, in the kidneys (Fig. 9.26), and in the skin. There may also be thrombocytopenia. In patients who recover, there is usually no residual disability.



Fig. 9.26 Fat embolism of glomerular capillaries in a case of caisson disease. The globules of fat (dissolved out in processing the tissue) have impacted in glomerular capillaries and caused great distension. $\times 170$. (Professor A. C. Lendrum.)

Air embolism. This occurs when air is aspirated into a severed vein, especially a large vein near the heart, but air may also enter the circulation in fatal amounts during blood transfusion if positive pressure is used without due care. The frequency and seriousness of the condition have probably been exaggerated. The air may produce effects in two ways. It may become mixed with the blood in the right ventricle, forming a froth which is not readily expelled and interferes with ventricular filling, or the bubbles of air may become arrested in the

pulmonary arterioles and lead to the mechanical effects of embolism. When air enters the circulation it is absorbed rapidly, and to produce serious results the sudden entrance of over 100 ml is usually necessary; less than this has provoked alarming symptoms, but recovery has occurred after as much as 300 ml. Injury to the spinal cord and bones can result from formation of bubbles of nitrogen in caisson disease, which develops in divers, etc., from too rapid decompression from a high atmospheric pressure (p. 779).

At necropsy, bubbles of gas are sometimes found in the blood, due to the action of the *Clostridium welchii* after death; this should not be mistaken for air embolism.

Parenchymal-cell embolism. In certain conditions special types of cells form emboli in the

pulmonary vessels, for example the megakaryocytes of the bone marrow in severe infections, the syncytial cells from the placenta, and hepatocytes after laceration of the liver. Such cellular emboli are without serious effect, the cells in all probability disintegrating. By contrast, the entry of **amniotic fluid** into the maternal circulation during prolonged or obstructed labour may cause serious effects in two ways. Firstly, it may produce extensive and sometimes fatal embolism of the pulmonary circulation by fetal squames, vernix and meconium. Secondly, it may bring about both widespread intravascular fibrin formation and activation of the plasminogen fibrinolytic system, with the result that there is severe hypofibrinogenaemia, and dangerous post-partum haemorrhage commonly results.

Local Ischaemia

Complete arterial occlusion

The term **ischaemic** is applied to tissue in which the blood flow has ceased (complete ischaemia) or is abnormally low (partial ischaemia). Ischaemia localised to an organ, a part of the body, or a patch of tissue, is usually due to obstruction to arterial blood flow.

By far the commonest and most important causes of complete arterial occlusion are thrombosis and embolism; other causes include proliferative changes in the intima of small arteries, and also arterial spasm as in Raynaud's disease or ergot poisoning.

When an artery is obstructed the result depends on the extent of collateral circulation, i.e. alternative vascular routes by which blood can reach the deprived tissue. The arterial anastomoses in the limbs are such that blockage of any one artery does not usually result in severe ischaemia provided that the other arteries are not seriously diseased. Similarly, there are effective collateral arteries in the integument and muscles of the trunk. In the internal organs, however, the anatomical arrangement of many of the vessels does not allow a sufficient anastomic supply, and severe ischaemia follows arterial occlusion. When an artery of a limb is suddenly obstructed in a

healthy subject, there is an immediate drop in the blood pressure beyond the obstruction, and the circulation is brought almost to a standstill; the arteries then contract and the part contains less blood than normally. Soon, however, the anastomotic arteries dilate and blood thus bypasses the obstruction to enter the vessels of the affected part, through which a flow of blood is gradually established and increased until ultimately it may approach normal. Thus in a healthy subject the femoral artery may be ligated without permanent damage resulting. The limb becomes cold and numb, and some time elapses before the pulse returns at the ankle; and it is much longer before complete muscular power is restored. The collateral vessels remain dilated and maintain the circulation, and in response to the sustained rise in blood flow there occurs a thickening of their walls, with increase of the muscular and elastic tissue corresponding with the enlarged lumen; in other words, the collateral vessels become permanently enlarged or hypertrophied. An outstanding example is seen in the rare congenital localised stenosis (*coarctation*) of the aorta beyond the arch, in which the vessels linking the arteries of the head and neck with those of the trunk and legs become enormously enlarged during life and supply

most of the blood to the lower part of the body.

The development of an efficient collateral circulation often depends on dilatation of healthy anastomotic arteries, and on a healthy heart. If, however, the collateral arteries are diseased, e.g. atheromatous, fibrosed or calcified, they are unlikely to dilate sufficiently to supply the necessary amount of blood to the ischaemic part, and a varying amount of necrosis will follow. Accordingly, in middle-aged or old people with atheroma, etc., blockage of the main artery of a limb, or even of a large branch, may be followed by death of the tissues supplied by the obstructed vessel, the condition of 'senile' gangrene resulting (p. 205). Multiple emboli in the arteries of the lower limbs (usually resulting from aortic atheroma) or spreading thrombosis, as in thromboangiitis obliterans (p. 378), may also lead to ischaemia and gangrene, even in young adults.

Infarction

Certain arteries of internal organs have imperfect anastomoses and their obstruction is always followed by serious results. Such arteries are called **end arteries**, and they may have no anastomosis, e.g. the splenic artery, or only capillary anastomosis, e.g. the branches of the renal artery, or arterial anastomosis insufficient to keep the part alive, e.g. the superior mesenteric artery. Obstruction of such vessels leads to ischaemia, usually sufficient to cause tissue necrosis. *The term infarct is applied to the altered area which has lost its blood supply, and use of the term implies that the tissue has undergone ischaemic necrosis.* During the process of infarction, the small blood vessels in the dying tissue may become engorged with blood from either retrograde flow or anastomosing small vessels. This engorgement, often accompanied by haemorrhages, results in the so-called **red** or **haemorrhagic** infarct. In other instances, little blood enters the dying tissues and a **pale** infarct develops. Pale infarcts occur in organs where there is little or no anastomosis, e.g. heart and kidneys; while in organs where there is some anastomosis, e.g. the intestine, and the lungs, red infarcts are found. Infarction means literally a stuffing-in, and was originally applied to the haemorrhagic type, which appeared stuffed with blood. When

pale infarcts were found to have a similar cause, the term was applied to them also.

Infarction is usually the result of acute occlusion of an artery by thrombosis or embolism. In the coronary arteries, atheroma with thrombosis is common, and is the usual cause of infarction of the myocardium; in the brain, thrombosis and embolism are both of importance but infarction can also result from hypotension; in the lungs, kidneys and spleen embolism is a commoner cause than thrombosis.

Features of infarcts in various sites. In the kidneys, spleen and lungs, the vascular arrangements are such that most infarcts are roughly wedge- or cone-shaped, the apex lying most deeply, in the vicinity of the occluded artery, and the infarct enlarging as it extends peripherally, the base being visible as a necrotic area on the surface of the organ (Figs. 9.27 and 9.30). The coronary arteries pass inwards from



Fig. 9.27 Two haemorrhagic infarcts of lung, seen on section as dark wedge-shaped areas, widening towards the pleural surface (*left*). Note the pulmonary artery occluded by thrombus (*arrow*) beyond the apex of the upper infarct. $\times 1.3$.

the epicardium, and accordingly myocardial infarcts involve especially the inner part of the wall, although commonly the whole thickness undergoes infarction.

In the **brain** a reduction in blood flow sufficiently severe to produce infarction is usually due to atheroma of the cerebral arteries or of major arteries in the neck that supply the brain, i.e. the internal carotid and vertebral arteries. The artery may be occluded by thrombus formed on an atheromatous patch or by an embolus, but stenosis alone, by severely impairing blood flow through the artery, may cause ischaemic damage in the brain. Indeed cerebral infarction may occur even when the arteries supplying the brain are normal. This sometimes results from a profound fall in cerebral blood flow due, for example, to an episode of severe hypotension. The sites of infarction of the brain depend on the cause of the ischaemia. When a major cerebral artery is blocked, infarction obviously occurs within the territory supplied by it, but blockage of the internal carotid or vertebral arteries, or a hypotensive episode, results in infarction in the so-called *boundary zones* (p. 743) at the margins of the territories supplied by the major cerebral arteries. Even when a major cerebral artery is completely occluded by thrombosis or embolism, there is considerable variation in the size of the infarct. This is due mainly to the efficiency of the potential collateral circulation through arteries on the surface of the brain and through the circle of Willis, both of which link the major cerebral arterial territories. A cerebral infarct may be pale or haemorrhagic. As the dead tissue soon breaks down and becomes soft, a cerebral infarct is often referred to as a *softening*. Thereafter, over a period of weeks or months, the necrotic tissue is gradually removed by phagocytes. The final result is a cystic shrunken area in the brain (see Figs. 21.25, 21.26, p. 744).

The central artery of the **retina** is an end-artery, and its obstruction causes retinal infarction, with loss of sight in the eye.

In the **heart**, obstruction of a coronary artery or a major branch gives rise to infarction of the ventricular myocardium; it is usually somewhat irregular in form and pale, but may show congestion and haemorrhage at the margin (Fig. 15.9, p. 404).

Obstruction of even a large branch of a pul-

monary artery does not always result in infarction. Experimentally-induced pulmonary emboli in otherwise healthy dogs do not usually cause infarction, some additional general impairment of pulmonary blood flow being required for infarction to result from the emboli, e.g. constriction of the pulmonary venous drainage. Similarly in man, a raised pulmonary venous pressure, due to mitral stenosis or heart failure, or to lung disease causing obliteration of pulmonary capillaries, predisposes to the development of infarction following pulmonary embolism. The subject is discussed more fully on pp. 458–60.

Pulmonary infarcts are typically wedge-shaped, with the base projecting slightly on the pleural surface (Fig. 9.27). They are firm and haemorrhagic (Fig. 9.28). In some instances, pulmonary arterial occlusion results in a wedge-shaped haemorrhagic patch without necrosis, and resolution may then occur, but when there is ischaemic necrosis, i.e. infarction, organisation and scarring follow in patients who survive for more than a few weeks.

Infarcts of the **spleen** are common and result usually from embolism. They are usually reddish at first and occasionally haemorrhagic, but soon become pale (Fig. 9.29) and yellow. In the **kidneys**, infarcts seen at necropsy are pale, with a deep red periphery due to congestion and

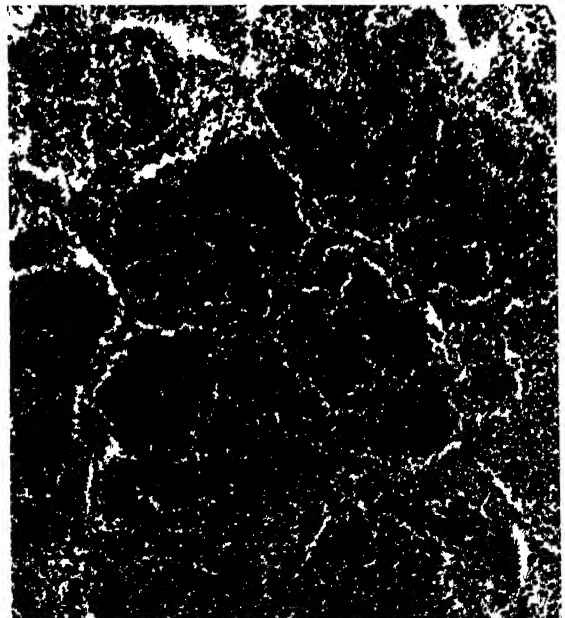


Fig. 9.28 Haemorrhagic infarct of lung, showing alveoli filled with red cells. $\times 115$.



Fig. 9.29 Pale infarct of the spleen.

haemorrhage. The pale portion involves chiefly the cortex, the affected part in the medulla being usually red and haemorrhagic (Fig. 9.30), but small infarcts may be haemorrhagic throughout. When an arterial branch in a kidney is blocked experimentally, the area supplied becomes at first swollen and red throughout owing to general congestion. Thereafter, the dying kidney cells take up water (p. 12), and their swelling expresses blood from the central part of the infarct, which thus becomes pale. At the periphery and in the medulla the hyperaemia and stasis persist, the ischaemic capillaries rupture, and haemorrhage occurs into the tissues.

Blocking of the superior mesenteric artery produces a haemorrhagic infarct of the intestine (Fig. 19.72, p. 645), which rapidly progresses to gangrene. Death usually results unless the infarcted intestine is removed surgically without undue delay. Obstruction of the inferior mesenteric artery may be without



Fig. 9.30 Infarct of kidney, showing pale necrotic centre with haemorrhagic margin. $\times 1.2$.

serious effect, but sometimes causes ischaemic colitis.

Haemorrhagic infarction of the intestine, following obstruction of the superior mesenteric artery, has been studied experimentally in dogs. When this artery is ligated there is at first an arrest of the intestinal blood flow, accompanied by a contraction of the muscular coats of the intestine. This soon passes off and blood flows into the vessels of the ischaemic segment from the arterial anastomoses, but is inadequate to restore the circulation. The capillaries and small veins become engorged and finally stasis occurs and there is diffuse haemorrhage into the wall of the intestine and its lumen. Complete deprivation of blood from 5–10 cm of the bowel was found to lead to haemorrhagic infarction.

In the liver, obstruction of a branch of the portal vein is not followed by infarction, owing to the supply of blood from the hepatic artery. The obstruction does, however, reduce the blood flow sufficiently to cause atrophy and loss of hepatic parenchymal cells, and the sinusoids become dilated, so that the lesion appears dark red and shrunken (Fig. 9.31). Obstruction of the hepatic artery or of its branches may result in infarction of the liver (p. 664).

'Venous infarction'. Obstruction of a vein is an uncommon cause of arrest of blood flow through the tissue it drains, partly because in most tissues there is sufficient anastomosis to maintain venous drainage, and partly because

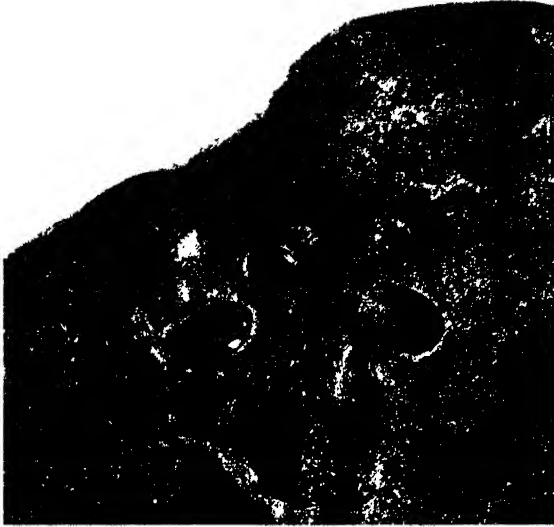


Fig. 9.31 A depressed red patch in the liver due to loss of parenchymal cells and sinusoidal congestion following thrombosis of a portal venous branch (not shown).

thrombosis of veins in internal organs is relatively rare, and emboli cannot, of course, impact in veins (except in portal venous systems, as in the liver). When venous infarction does occur the infarct is intensely engorged, oedematous and haemorrhagic. Marantic thrombosis of the superior longitudinal sinus sometimes occurs in severely debilitated children: the engorged cerebral cortical veins may rupture (Fig. 9.32), and there may be patches of haemorrhagic infarction of the cortex. Thrombosis of the mesenteric veins extending down to the smaller tributaries causes infarction of the intestine, which progresses to gangrene. Venous infarction is also seen occasionally in the liver as a result of extension of hepatic cancer into the hepatic veins. The best examples of venous infarction are, however, seen in the adrenals, which have several arteries but drain through a single large vein.

The susceptibility of tissues to ischaemia. The extent of infarction is usually less than that of the tissue supplied by the occluded artery, collateral circulation supplying the tissue at the periphery of the area. The size of the infarct resulting from occlusion of a particular artery may thus vary greatly, depending on whether the collateral arteries are healthy and capable of dilatation. The extent of necrosis is determined also by the capacity of the tissue to withstand ischaemia. As a general rule, *the*



Fig. 9.32 Thrombosis of the superior longitudinal sinus (shown below), resulting in intense engorgement of the cerebral cortical veins and haemorrhage over the frontal lobe.

parenchymal cells of the internal organs, which operate at a high metabolic rate, are relatively susceptible to ischaemia, whereas the supporting tissues—connective and fatty tissue and bone, are much less susceptible. The neurones of the central nervous system are perhaps the most susceptible cells of all, and cannot withstand deprivation of blood supply for more than a very few minutes. Glial cells are somewhat less demanding in their requirements, and accordingly at the margin of a brain infarct there is a zone in which partial ischaemia is followed by restoration of the circulation by collaterals; this results in death of the neurones, while the glial cells persist and undergo reactive proliferation. Hepatic parenchymal cells are also highly susceptible to ischaemia and, as described above, thrombosis of a portal venous branch is commonly followed by atrophy and loss of liver cells with survival and dilatation of the sinusoids. The renal tubular epithelium has also a low resistance to ischaemia, and while in the central part of a recent renal infarct all the cells are dead, at the periphery there is a zone in which the glomeruli and intertubular capillaries have survived while the tubular epithelium has died.

Changes following infarction. The autolytic changes which follow infarction have been described on pp. 11–13. Depending on the nature of the tissue, and whether or not it is oedematous or haemorrhagic, the infarct may remain firm (*coagulative necrosis*) or soften (*colliquative necrosis*).

From an early stage of infarction, products

of the breakdown of ischaemic or dead cells at the edge of the infarct diffuse into the surrounding tissue and promote a mild **acute inflammatory reaction**, with exudation of fluid from the vessels and migration of neutrophil polymorphs into the peripheral dead tissue. This, together with ischaemia of their walls, accounts for the dilatation of the small vessels and haemorrhage at the margin of the infarct. The acute reaction soon passes off, and the emigrated polymorphs die. The dead tissue stimulates a reaction similar to that around a foreign body: within a few days a zone of vascular **granulation tissue** forms around the infarct and the dead tissue is gradually organised. Macrophages migrate into and digest it from the periphery inwards; they are accompanied by new capillary buds and fibroblasts, so that the granulation tissue extends centrally, and as it matures into fibrous tissue the infarct is gradually converted to a **scar**. Extravasated red cells soon lose their outlines, and the pigment is slowly absorbed, although haemorrhagic infarcts, e.g. in the lungs, remain brown for a long time, and macrophages containing haemosiderin may long remain in and around the scar tissue. Loss of parenchymal cells and contraction of the fibrous tissue results in shrinkage; thus an old infarct of the myocardium presents the appearance of fibrosis and thinning of the ventricular wall (Fig. 15.11, p. 405). In solid organs, old scarred infarcts result in surface depressions, and when they are multiple, e.g. from repeated emboli in the branches of the renal arteries, the surfaces may be puckered and deformed by the scarring.

Septic infarcts. Multiple infarcts due to emboli containing pyogenic bacteria are an essential feature of pyaemia (p. 201) and thus a complication of acute bacterial endocarditis, acute osteomyelitis, carbuncle, etc. The bacteria may extend into, and multiply in, the dead tissue, and this results in an acute inflammatory reaction in the tissue around the margin of the infarct; thus the central dead tissue becomes surrounded by a ring of suppuration (Fig. 9.33).

The effects of infarction. The effects of infarction on function depend largely on the location and size of the infarct. In organs such as the kidneys, which have a large functional reserve, extensive or multiple infarctions of



Fig. 9.33 Suppuration at the margin of a septic infarct of the heart. The necrotic myocardium (*right*) is becoming separated from the adjacent living tissue (*left*) by a purulent exudate. $\times 175$.

both kidneys are necessary to bring about any serious disturbance of function, and serious impairment of liver function also requires very extensive infarction. By contrast, single infarcts of the myocardium are commonly sufficiently large to reduce seriously the functional reserve of the heart, and cause heart failure; infarcts involving the conducting system of the heart may cause heart block, and occlusion of a coronary arterial branch not uncommonly causes death from ventricular fibrillation before infarction has become apparent. Infarcts of the brain are a major cause of serious dysfunction, and even a small one involving the internal capsule is followed by hemiplegia. As already explained, infarction of lung tissue tends to occur especially in association with embarrassment of the pulmonary circulation, and for this reason recent pulmonary infarcts are quite commonly observed at necropsy of patients dying of heart failure.

The effects of infarcts are considered more fully later, in the systematic chapters.

Partial arterial obstruction

Chronic narrowing of the lumen of arteries is very common, and is usually caused by atheroma (p. 363). It brings about the serious effect of ischaemic atrophy of specialised cells with accompanying overgrowth of fibrous tissue, for example in the myocardium and the kidneys. Atheromatous narrowing of the arteries which supply the brain predisposes to focal loss of neurons or to actual infarction (p. 742): these events are particularly liable to occur during hypotensive episodes and no doubt contribute to intellectual deterioration in old age. Prevalence of atheroma in the elderly is an important cause of senile mental changes. Multiple or

extensive atheromatous narrowing of the lumen is common in the arteries of the lower limbs, and the resulting chronic ischaemia brings about various trophic changes, and also limping, and cramp-like ischaemic pain, induced by walking (*intermittent claudication*). Narrowing of the smallest arteries and arterioles—*arteriosclerosis*—occurs commonly in the abdominal viscera and central nervous system as an ageing effect, and results particularly from arterial hypertension: it is usually most severe in the afferent arterioles of the glomeruli, where it brings about glomerular sclerosis. These regional changes are, however, more appropriately considered in relation to the various systems and organs.

Disturbances of Water and Salt Balance

Water and salt deficiency

The water content of the average male body, estimated by the deuterium method, is about 62 per cent, and that of the female about 52 per cent, the sex difference being accounted for by the higher fat content in females. A man weighing 70 kg contains about 42 litres and this is distributed as 30 litres of *intracellular* water and 12 litres of *extracellular* water; the latter is subdivided into about 3 litres of *intravascular* fluid, the plasma, and about 9 litres of *interstitial* fluid which is distinguished from the intravascular and intracellular fluids by its very low protein content. The extracellular fluids contain practically all the sodium (except for that associated with collagen and that forming part of bone mineral), balanced chiefly by chloride and bicarbonate ions, whereas the intracellular fluid is almost devoid of sodium and chloride, its proteinate, sulphate and phosphate anions being balanced by potassium and magnesium. It is essential that the interstitial fluid should remain isotonic with the intravascular and intracellular fluids, and it contains a higher concentration of electrolytes which balances the colloid osmotic pressure of their proteins. Reductions in the water and salt content of the body are generally associated, but disproportionate depletion of either water or salt causes disturbances of the normal equilibrium which

require different treatment. Deficiency of water tends to cause hypertonicity of the extracellular fluids so that water is withdrawn from the cells, which thus share in *primary dehydration*. Conversely, in relative salt depletion the extracellular fluids tend to become hypotonic, but this effect is minimised partly by increased renal excretion of water and partly by diffusion of water from the interstitial fluid into the cells with maintenance of isotonicity. Thus in salt deficiency the extracellular fluids are reduced in volume, but the administration of water or glucose solution without salt is actually harmful as it merely dilutes further the extracellular fluids and increases the diffusion of water into cells. *It is curious that whereas the need for water is normally indicated by thirst, in man there appears to be no urgent warning sensation when salt is lacking.*

Dehydration may be brought about in various ways and in minor degrees is very common. In hospital patients it is seen most often as a result of insufficient intake owing to physical weakness, coma and pyrexia. The urine is reduced in volume (500 ml) and is highly concentrated, the specific gravity rising to 1.040 or more. The plasma levels of Na^+ , Cl^- and urea increase, probably as the result of diminished renal filtration, although the plasma volume is maintained relatively well by withdrawal of intracellular water and by active retention of Na^+

and excretion of K^+ under the influence of the renin-angiotensin-aldosterone system (p. 258) which is stimulated by the diminished blood volume—the so-called reaction of dehydration. More severe dehydration occurs under exceptional conditions, e.g. in people shipwrecked or lost in the desert, and then the deficiency of body water may ultimately reach over 12 per cent of body weight and amount to nearly 10 litres. Death is thought to be due to rise in the osmotic pressure of the cells. In children the ratio of body surface to weight is higher than in adults so that cutaneous losses of water are proportionately greater; also children cannot produce such a high concentration of urine as can adults. As a result, lack of fluid has a more severe effect in infants and young children than in adults.

Salt depletion is a commoner cause of serious effects than is water depletion, and also is more liable to remain unrecognised. Excessive loss of sodium chloride from the body occurs in various conditions and is commonly only one factor in complex fluid and electrolyte disturbances. Pure loss of salt results from excessive sweating when water is consumed freely, e.g. in the tropics or when working in a very hot atmosphere. It gives rise to a state of 'heat exhaustion' which necessitates the administration of large amounts of salt as well as of water, the consumption of water alone being liable to produce severe cramps. Clinically, vomiting and diarrhoea are the most important causes of combined water and salt depletion: the former is complicated by alkalosis due to loss of H^+ , and the latter by acidosis from loss of the alkaline secretions of the small intestine. If water alone is restored the picture of pure salt depletion follows: lowering of the osmotic pressure of the extracellular fluid leads to renal excretion of water and to increased osmotic absorption of water by the tissue cells. In consequence, there is severe depletion of the extracellular fluid and circulatory collapse (*shock*) soon supervenes. The effects of this *secondary extracellular dehydration* are actually more serious than those of the disturbed acid-base balance which may develop from disproportionate loss of sodium or chloride ions, though the latter condition at one time received more attention. The symptoms of salt depletion when water is consumed freely include lassitude, weakness, giddiness, fainting attacks and

cramps; also anorexia, nausea and vomiting occur and tend to set up a vicious circle. Marked loss of weight and mental confusion may occur also. The plasma concentration of sodium, normally about 137–148 mmol/litre, falls to 130–120 mmol/litre or less. The chloride and bicarbonate concentrations are also reduced *in toto* but their ratio varies with the presence of complicating acidosis or alkalosis. The blood is concentrated, with a rise in haemoglobin, haematocrit value and in plasma proteins. The urine contains little or no sodium or chloride except when the salt depletion is due to renal loss, as in Addison's disease or diabetic ketosis. The blood urea rises, often to over 17 mmol/litre (100 mg per 100 ml), owing mainly to reduced renal blood flow and diminution in the volume of glomerular filtrate—*pre-renal uraemia* (p. 847).

Combined deficiency of water and salt is more common clinically than of either separately. Vomiting and diarrhoea are probably its most frequent cause. If water is ingested and retained, salt deficiency will predominate, as described above, but without fluid intake water loss exceeds salt loss. In such combined deficiency, the extracellular fluid therefore tends to become hypertonic and consequently fluid is withdrawn from the cells; this leads to symptoms of salt depletion (see above) and unless corrected may cause acute circulatory failure. The rise in blood urea often leads to the erroneous diagnosis of uraemia due to renal failure, but the administration of water and salt in adequate amounts may completely relieve the symptoms.

Regulation of the water content of the blood and urine is normally carried out by the kidneys, which in turn are controlled largely by secretion of antidiuretic hormone by the neurohypophysis: this regulates resorption of water in the distal renal tubule. The neurohypophysis is so highly sensitive to the osmotic influence of sodium chloride that an alteration of 1 per cent in the osmotic pressure of the arterial blood can bring about a tenfold variation in the excretion of water, and the osmotic pressure of the extracellular fluids is thus very precisely controlled. Failure of this mechanism is seen in diabetes insipidus (p. 1015), in which there is intense polyuria approaching maximum water excretion. An analogous situation in respect of excessive salt excretion results from

failure of the secretion of adequate amounts of aldosterone by the adrenal cortex, e.g. in Addison's disease, in which the cortex is largely destroyed. Uncontrolled sodium loss in the urine leads to fall of the plasma sodium to far below the level at which it normally ceases to be excreted. In consequence serious depletion of the body's store of sodium is brought about and this, if uncorrected, contributes greatly to the severe crises of Addison's disease and the tendency to acute circulatory collapse (p. 1044). Other hormones also play minor parts in the regulation of water and salt excretion, e.g. ovarian hormones can cause a distinct retention of water, as is seen in the late phase of the menstrual cycle and in pregnancy.

The pathology of generalised oedema has to be viewed against this background of water and salt balance. Maintenance of osmotic equilibrium is more important for life and is therefore regulated more exactly than the total volume of fluid in the body or within any of its compartments. Most importance was formerly attached to the chloride anion, but it is now recognised that the sodium cation is even more significant in regulating the amount of body fluid in the extracellular compartment of the tissues, and that sodium is intimately concerned in the pathogenesis of oedema.

Water and salt retention: oedema

Oedema is an abnormal increase in the amount of interstitial fluid. It may be localised, e.g. in an organ, limb, etc., or more generalised. In generalised oedema there is usually accumulation of fluid also in the serous cavities (hydrothorax, ascites, etc.). When oedema affects the skin and subcutaneous tissue swelling may be obvious, and momentary pressure will produce a depression ('pitting') which disappears in a few seconds as the oedema fluid returns to the tissue.

Control of interstitial fluid. The total exchange between the plasma and interstitial fluid is probably of the order of 7000 litres of fluid daily. In individual tissues exchange fluctuates with the physiological changes in blood flow. It is generally accepted that the interchange of fluid between the capillaries and venules and the tissue spaces can be explained on a physical basis, the distribution of fluid within and outside the vessels being regulated mainly by a

balance of the two processes of filtration and osmosis (pp. 48–9). The permeability of the capillary walls varies in different regions and also under different conditions of physiological activity in any one region, but the filtrate in all situations normally contains at least a small amount of protein, probably not exceeding 0.5 per cent in the more permeable areas such as the liver, and less than 0.1 per cent in the less permeable areas such as the limbs. In the normal exchange of interstitial fluid between vessels and tissue spaces most of the filtrate is returned to the circulation by the veins and only a small amount by the lymphatics, but most of the protein escaping from the vessels is carried away in the lymphatic fluid, the protein content of which therefore varies greatly in different parts of the body, depending on the permeability of the capillaries in the area drained: for example, the hepatic lymph is very rich in protein (3–5 per cent).

Water retention. It is important to appreciate that, regardless of its cause, generalised oedema represents retention of water and does not arise from a mere redistribution of the body fluids. In an adult, an increase of weight of about 5 kg invariably precedes the appearance of clinically recognisable generalised oedema, a fact utilised in the attention paid to the weight during pregnancy. Indeed generalised oedema can be regarded as a method of disposing of excess fluid which cannot be discharged by the usual channels, in order to regulate the blood volume. The body appears to tolerate badly an increase in the volume of the intravascular fluid; the excess is shunted into the interstitial spaces where its presence requires the simultaneous retention of a sufficient quantity of electrolytes, chiefly salt, to equalise the osmotic pressure of this fluid with that of the cells and of the plasma. The osmotic effect of the intracellular and plasma proteins is balanced by a higher concentration of electrolytes—chiefly salt—in the interstitial fluid. It is unlikely that increase of capillary permeability to macromolecules plays any major part in the common forms of generalised oedema, for the protein content of oedema fluid is not sufficiently high to suggest this possibility. Also, there is no gross fall in the blood volume, as might be expected if exudation of protein-rich plasma fluid was an important factor. In rare cases, cyclical oedema has been accom-

panied by hypovolaemia, and it has been suggested that the oedema of hypothermia may be related to an increase of factors such as bradykinin, which increase capillary permeability.

Local oedema

Active hyperaemia: inflammatory oedema. Active hyperaemia occurs in acute inflammation, in which the exudation of protein-rich fluid from the capillaries and venules gives rise to inflammatory oedema: as indicated in Chapter 3, major factors in the production of inflammatory oedema are increased hydrostatic pressure in the small vessels with dilatation and increased vascular permeability.

Active hyperaemia of lesser degree occurs also under physiological conditions, for example in the skeletal muscles during exercise, in the gastro-intestinal tract during digestion, and in the skin as an important mechanism of heat loss: the increase in interstitial fluid resulting from such physiological hyperaemia is, however, removed by the lymphatics and oedema does not result.

Oedema is a prominent feature of some types of **hypersensitivity reactions**, for example in hay fever, urticaria, the Arthus and delayed hypersensitivity reactions. These are all described in Chapter 6, and it is sufficient to state here that the oedema is of inflammatory nature, due to active hyperaemia and increased vascular permeability.

Urticaria consists of erythema, itching and wealing (which is sharply localised oedema) of the skin: it is very common, but occurs usually in mild form, with only occasional transient attacks. In some instances, however, attacks are frequent, severe, or more persistent. Although there is often a clear association with eating a particular food or taking a drug (especially aspirin), there is little firm evidence of an immunological hypersensitivity basis in most cases, and the underlying nature of the condition is usually unknown. Histamine antagonists are often beneficial, but in therapeutic dosage these agents have various other effects in addition to anti-histamine activity. Except when it occurs as part of an anaphylactic attack, urticaria is seldom dangerous.

Hereditary angio-oedema. This rare condition is characterised by attacks of acute localised oedema, most often affecting the skin of the

face and trunk, but sometimes the larynx: acute abdominal pain, vomiting and diarrhoea can also occur as a result of oedema of a segment of gut, and some patients have undergone several abdominal operations. Death from laryngeal oedema is common in some affected families.

The condition is due to a genetically-determined (autosomal dominant) abnormality of an inhibitor of the activated first component of complement (C1-INH). In some cases there is insufficient inhibitor, in others it is qualitatively abnormal. The oedema is possibly due mainly to a kinin-like fraction of activated C2, and transfusion of fresh normal plasma (which contains the inhibitor) is of temporary value in both prevention and treatment of attacks.

Oedema may occur in severe cases of **zoster** (shingles) and is apparently a trophic effect due to inflammatory change in the posterior root ganglia. If the nerve lesion is unilateral, as it usually is, the oedema stops short in the mid-line of the body.

Local venous congestion and oedema. In a healthy animal, acute venous congestion produced by ligation of a large venous trunk does not usually lead to oedema, although there is an increased filtration of water and electrolytes owing to the heightened capillary pressure, and also an increase in the amount of protein leaving the vessels. Consequently there is increased flow of lymph containing a lowered concentration, but increased amount, of protein, and oedema does not usually develop. If, however, along with the ligation of the vein the vasomotor nerves supplying the part are cut, the intracapillary pressure is still further increased and localised oedema follows. Similarly, the application of an elastic band to a limb may merely produce venous congestion with increased lymph flow unless the band is tightened sufficiently to prevent the flow of lymph from the part, when oedema will result. These findings indicate that some other factor in addition to acute venous congestion is usually necessary for the production of oedema. In clinical cases, however, local venous congestion often lasts much longer than in the experimental animal, and this may possibly explain the common occurrence of oedema.

In acute venous obstruction there must be sufficient anastomotic drainage of venous blood to permit the circulation to continue; otherwise

stasis, thrombosis and haemorrhagic infarction would follow as is seen in mesenteric venous thrombosis. After a time readjustment of the circulation occurs and arterial inflow diminishes. This leads to a reduction in tissue perfusion until, in time, the collateral circulation increases sufficiently to re-establish normal drainage. Until that stage is reached, the combination of venous congestion, tissue hypoxia and accumulation of metabolites may, by increasing capillary hydrostatic pressure and permeability, result in local oedema.

The oedema of chronic lymphatic obstruction, e.g. that produced by cancer, chronic inflammation, radiotherapy, filariasis, etc. (p. 393), is usually of the non-pitting type, i.e. the swollen tissues do not yield readily to pressure. A characteristic feature of chronic lymphatic oedema is the development of elephantiasis due to overgrowth of the connective tissue in the skin and subcutaneous tissue. Since the plasma protein normally present in the interstitial fluid is returned to the blood by the lymphatics, chronic lymphatic obstruction results in the accumulation of protein in the tissues while most of the water and electrolytes are taken up by the venules as usual. This accumulated protein may, in some unknown way, be responsible for stimulating the connective tissue cells to increased production of collagen.

Lymphatic oedema of the legs also occurs as a primary condition which is sometimes hereditary (Milroy's disease) and is believed to be due to a congenital abnormality of the lymphatics. The legs become permanently thickened.

General oedema

Cardiac oedema is apt to develop at a late stage in cases of **right ventricular failure** with long-standing systemic venous congestion. It appears first in the most dependent parts of the body and gradually extends upwards. Thus it is usually noticed first round the ankles, and pitting may be elicited by pressure over the lower end of the tibia. When the condition is advanced, the limbs become greatly swollen, the skin is tense and vesicles may form. Accumulation of fluid may occur also in the serous cavities.

As indicated above, increased transudation from congested, dilated capillaries is not sufficient to produce oedema experimentally be-

cause the excess fluid is removed by the lymphatics. When, however, heart failure becomes severe, the diminution in cardiac output adversely affects renal function which depends upon normal renal blood flow. The kidneys can compensate to some extent for reduced blood supply by increasing the proportion of fluid filtered off in the glomeruli; this is probably mediated by increased tone in the efferent arterioles. The volume of urine is reduced and it is highly concentrated, indicating that there is excessive tubular re-absorption of water. The mechanism of this excessive re-absorption is not fully understood, but the reduced renal blood flow may stimulate the juxta-glomerular cells to secrete excess of renin, and this in turn will enhance the secretion of aldosterone by the adrenal cortex, with consequent re-absorption of sodium by the renal tubules. The effect of sodium retention is to stimulate secretion of anti-diuretic hormone by the neurohypophysis, and so more water is re-absorbed in the renal collecting tubules. This mechanism has been demonstrated to play a role in some, but not all, cases of cardiac oedema (p. 258). The stimulus to this secondary aldosteronism is not fully understood, as it occurs among different types of heart failure, both in low output and in high output types. The great increase in body weight confirms the enormous amount of fluid retained in the oedematous tissues in some cardiac cases, and the importance of water and salt retention is shown by the effect of diuretics in diminishing the oedema. Reduction in the intake of sodium chloride in the diet has sometimes a markedly diuretic effect, water being eliminated with preservation of the isotonic state of the oedema fluid.

Other factors may play a part in the genesis of cardiac oedema, e.g. the accumulation in the tissues of waste products which by their osmotic action will tend to attract more water from the blood. There is also evidence that chronic hypoxia increases capillary permeability, although this view is not supported by the protein content of the oedema fluid, which is usually about 0.5 per cent or less. The fundamental cause of cardiac oedema appears, however, to lie in the faulty elimination of fluid consequent upon the deranged renal circulation. The distribution of the retained fluid in the tissues is determined by gravity because, with the reduction in cardiac power, the circulation is unable

to absorb the tissue fluid and return it to the right heart against the hydrostatic pressure of the column of venous blood in the dependent parts. The distribution of oedema fluid is influenced also by the degree to which different tissues can be distended without a significant rise in tissue pressure.

In **failure of the left ventricle** of the heart, venous congestion occurs mainly in the lungs so long as the right ventricle continues to beat forcibly: pulmonary oedema may then develop without generalised oedema (p. 398).

Renal oedema. Generalised oedema occurs in various diseases which affect the glomeruli including some types of glomerulonephritis, and also in acute renal failure due to injury to the renal tubules. The pathological changes in these conditions are described in Chapter 22. However, an understanding of the factors likely to be involved in the production of the various types of renal oedema depends not so much on a knowledge of the detailed structural changes but rather on the associated functional disturbances. Accordingly, renal diseases which give rise to oedema may be classified into the following three groups.

(1) Conditions in which all the glomeruli are affected, with reduction in renal blood flow and in glomerular filtration. This group is exemplified by *acute diffuse glomerulonephritis* and *rapidly progressive glomerulonephritis*. There is usually a rise in blood pressure and blood urea level, and production of a diminished amount of concentrated urine containing moderate amounts of protein. The oedema in these conditions is not influenced by gravity to the same extent as is cardiac oedema and is often noticed first in the loose connective tissues, e.g. of the eyelids and face: in ambulant patients, however, gravity is seen to have some effect. The protein content of the oedema fluid is usually less than 0.5 per cent and the oedema therefore cannot be attributed to increased capillary permeability. Also the proteinuria is usually only moderate and the loss does not result in any significant reduction in the levels of the plasma proteins. The blood volume is normal or increased, and the oedema seems likely to be due to excessive re-absorption of salt and water in the renal tubules. The factors responsible for this excessive re-absorption are not, however, clearly defined. The renin-angiotensin-aldosterone system

may be implicated, but even this is uncertain.

(2) *The nephrotic syndrome.* In some renal diseases there is persistent and heavy loss of plasma proteins, particularly albumin, in the urine: when this exceeds about 10 g daily, the plasma albumin level falls considerably and this is accompanied by generalised oedema which often becomes very severe. This condition is known as the nephrotic syndrome. As in other types of renal oedema, the distribution of the tissue fluid is not so dependent on gravity as in cardiac oedema. In patients with nephrotic syndrome, the blood pressure is often not raised and there is commonly no rise in the blood urea, indicating that renal blood flow and glomerular filtration rates are approximately normal. The nephrotic syndrome may arise in a large number of conditions: in some it is regularly present, for example in *glomerulonephritis of minimal-change* and *membranous* types (q.v.). It is a common result of *amyloid disease* involving the glomeruli; it occasionally complicates other types of glomerulonephritis and the glomerular lesions of diabetes mellitus and various other diseases. *In all these conditions, its development is dependent on excessive loss of plasma albumin into the glomerular filtrate.*

Glomerular leakage of protein exhibits a molecular sieving effect, the amount of plasma albumin which escapes being disproportionately great because of its relatively small molecular size. Also because of its small size and its relatively high concentration in the plasma, albumin is the protein mainly responsible for the osmotic pressure of the plasma, and consequently, in states of severe hypoalbuminaemia, the amount of fluid leaving the capillaries and venules throughout the body greatly exceeds the amount drawn back into them by osmosis. Accordingly, the plasma volume tends to fall and this brings into play the renin-angiotensin-aldosterone mechanism, resulting in increased re-absorption of sodium and water from the renal tubules: this tends, in turn, to dilute the plasma protein still further and so transudation into the tissues remains excessive; a vicious circle is set up and continues to operate so long as gross albuminuria persists. As would be expected, the oedema fluid in the nephrotic syndrome has a very low protein content, and there is no evidence of general in-

creased capillary permeability for macromolecules.

The importance of protein loss and hypoalbuminaemia is confirmed by the appearance of similar gross oedema in protein-losing enteropathy (p. 639) in which the kidneys are normal and there is gross loss of plasma protein into the gut. The participation of the renin-angiotensin-aldosterone mechanism is demonstrated by the very high plasma levels of aldosterone found in the nephrotic syndrome.

(3) In *acute tubular injury*, the tubules lose their capacity for selective re-absorption and concentration of the glomerular filtrate. Consequently, most of the filtrate is re-absorbed and the small amount of urine produced approximates in its composition to a protein-free filtrate of plasma. There is retention of water and electrolytes and a progressive rise in blood urea. Apart from loss by sweating, vomiting, etc., most of the fluid taken by mouth is retained in the body and unless it is seriously restricted, gross oedema develops.

Acute tubular injury may result from shock or certain chemical poisons (p. 848). In some cases of acute renal failure following shock, the tubules show no convincing evidence of necrosis in a renal biopsy, and it now appears that hyperactivity of the renin-angiotensin-aldosterone system may be of importance (p. 850).

In the various forms of renal disease which are complicated by arterial hypertension, cardiac failure is liable to develop with consequent generalised or pulmonary oedema.

Nutritional oedema

Generalised oedema may clearly be caused by malnutrition. Protein insufficiency seems to be the main factor, and the extreme example is termed kwashiorkor (p. 667). Examination of the plasma shows a marked fall in glucose, lipids and proteins, the last being sometimes reduced to half the normal. It seems likely that a fall in osmotic pressure of the plasma is the most important factor in the production of the oedema, but no strict parallelism has been found, some cases failing to become oedematous in spite of severe depletion of serum albumin, while others show gross oedema with plasma protein levels within normal limits; also the oedema may disappear before there is any significant rise in the colloid osmotic pressure of the plasma. Nutritional oedema is commonly associated with xerophthalmia, a condition in which opacity with ulceration of the cornea occurs, as a result of deficiency in fat-soluble

vitamin A; possibly lack of the vitamin B complex is also concerned, and the 'wet' (oedematous) form of beri-beri is perhaps related. A similar form of oedema has been observed in infants when there has been excess of carbohydrates in the diet with marked deficiency in other foodstuffs. In all such examples of nutritional oedema, the problem is a complex one, and the relative importance of the various factors outlined above varies in individual cases.

Oedema may occur in patients with chronic wasting diseases, e.g. cancer, tuberculosis, etc., and is due mainly to cardiac failure, although fall in the plasma proteins is a contributory factor in some cases.

Pulmonary oedema

The osmotic pressure of the plasma (25 mm Hg) is substantially greater than the normal hydrostatic pressure in the pulmonary capillaries (8–10 mm Hg). Consequently, the development of oedema of the lungs usually requires a considerable rise in the hydrostatic pressure. As elsewhere, this occurs, together with increased vascular permeability, in acute inflammatory lesions, and inflammatory oedema is pronounced in severe influenza and lobar pneumonia, etc.

Pulmonary oedema can be produced readily in healthy dogs by interfering with the flow of pulmonary venous blood, for example by compressing the left atrium or ventricle, or constricting the aorta. Similarly, in man, it occurs in left ventricular failure, as in some cases of myocardial infarction and in systemic hypertension. In this latter condition, acute pulmonary oedema comes on especially when the patient is lying down, probably due to improved venous return from the legs, and perhaps also to increase in the blood volume by re-absorption of oedema fluid from the legs when recumbent. The attack is usually relieved by sitting up. Chronic pulmonary congestion, as, for example, in stenosis of the mitral valve, is not alone sufficient to produce pulmonary oedema in man. This is probably because reflex increase in tone of the hypertrophied pulmonary arterioles protects the pulmonary capillary bed from excessive rise in pressure. However, the situation is precarious, and pulmonary oedema is prone to result from physical exertion or other factors which increase the pulmonary blood flow. Chronic pulmonary oedema may occur as part of generalised renal oedema, particularly when there is, in addition,

systemic hypertension, as in acute glomerulonephritis. Another important cause is overloading of the circulation by rapid transfusion of blood to patients with severe anaemia. Finally, pulmonary oedema occurs in some cases of increased intracranial pressure, most commonly in head injury or intracranial haemorrhage.

Apart from the above causes, oedema of the posterobasal parts of the lungs is a very common finding at necropsy, particularly in old people and where death is due to a toxic condition or has been preceded by coma. The oedema fluid is very prone to become infected by a mixture of bacteria, usually of low virulence, producing *hypostatic pneumonia* which, if untreated, is likely to be the immediate cause of death.

Depending on the causal factors, pulmonary oedema may be confined within the alveolar walls, i.e. interstitial oedema, or the fluid may pour into the alveolar spaces. The factors concerned are considered on pp. 451–4.

The renin–angiotensin–aldosterone system

Renin is an enzyme, stored and probably formed in the renal juxta-glomerular apparatus; it is present in high concentration in the cytoplasmic granules of cells of the wall of the terminal part of the afferent glomerular arterioles. The adjacent macula densa, a plaque of specialised epithelial lining cells in the wall of the distal convoluted tubule, is probably a sensory device, regulating the release of renin in response to changes in the composition of the fluid in the tubular lumen (Fig. 9.34).

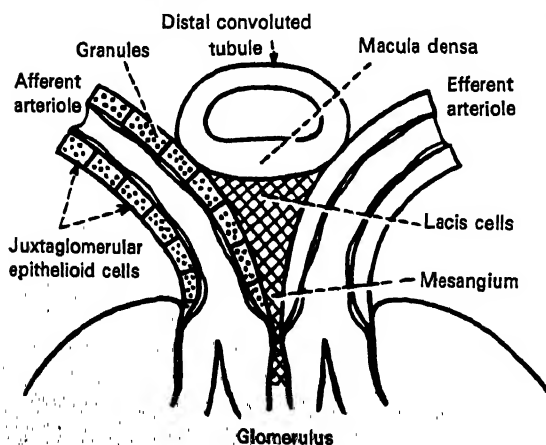


Fig. 9.34 Diagram of the juxta-glomerular apparatus.

Renin-substrate (angiotensinogen) is present in the α_2 -globulin fraction of plasma and also in renal lymph.

The initial product of the action of renin on its substrate is an inactive decapeptide, angiotensin I, which is converted in the circulation to the active octapeptide, angiotensin II. This conversion takes place largely in the pulmonary circulation, but it occurs also in the kidneys and this is of considerable importance in considering the possible direct renal actions of angiotensin.

Renin is normally present in higher concentration in renal lymph than in renal venous plasma, but because of the much higher rate of renal plasma flow, secretion into the renal vein is greater than that into lymph.

Effects of renin. Renin, by way of angiotensin, has three principal actions:

- (a) Aldosterone-stimulating.
- (b) Pressor, mediated mainly by peripheral vasoconstriction.
- (c) A direct renal effect, modifying urinary output of water and electrolytes.

Other actions, which hitherto have been less fully studied, are the central stimulation of thirst, release of catecholamines and vasopressin.

The relative dominance of the three principal actions mentioned above is much modified by the prevailing sodium status. Sodium deprivation, for example, enhances the aldosterone-stimulating effect, while minimising the pressor action, so that a marked rise in circulating renin, angiotensin II, and aldosterone occurs with little or no increase in arterial blood pressure.

The renal effects of administered angiotensin vary widely according to the dosage, the prevailing sodium status, arterial pressure, and species. At most doses which can safely be given to normal man, angiotensin reduces renal excretion of sodium and water, and this effect is enhanced by severe sodium depletion, as in untreated Addison's disease. By contrast, in hypertension, irrespective of aetiology, and also in hepatic cirrhosis with ascites, angiotensin usually increases water and sodium loss.

Secondary hyperaldosteronism. The renin–angiotensin–aldosterone system is stimulated, and high circulating levels of all three components may be found, in sodium depletion, whether due to dietary sodium restriction, sodium-losing renal disease, diuretics or purgatives: haemorrhage produces a similar response. Because in these situations the increase in aldosterone is thought to be a consequence of a rise in renin, these are regarded as examples of 'secondary' hyperaldosteronism.

Secondary hyperaldosteronism develops in some, but by no means all, cases of *untreated congestive heart failure*, in *hepatic cirrhosis with ascites*, and in

the nephrotic syndrome. It may seem paradoxical that patients with these oedematous states, with their retention of sodium and water, should react as though sodium-deprived. The explanation lies probably in that the excess sodium is principally extravascular, and thus not capable of recognition by the kidney. The kidney therefore responds as in sodium deprivation; hence plasma renin and angiotensin, and in consequence, aldosterone, are elevated.

Renal artery stenosis is another instance in which the kidney probably receives a stimulus to increased renin release which is inappropriate to the overall requirements of the body. The old belief that renal artery constriction, by leading to increased circulating renin and angiotensin, is simply and directly responsible for hypertension *via* the pressor effect of angiotensin (p. 375) is a considerable oversimplification. It is clear, however, that in many cases of severe renal artery stenosis with hypertension, both renin and aldosterone are increased. A similar mechanism—possibly multiple intrarenal arterial lesions—may be the cause of the secondary hyperaldosteronism which often accompanies the malignant phase of hypertension, irrespective of aetiology. In advanced chronic renal disease with renal failure, occasionally such severe elevation of renin and aldosterone levels may occur that hypertension cannot be controlled until both diseased kidneys have been excised. This could well be an instance where sufficient angiotensin is circulating to have a direct pressor effect. However, angiotensin II also raises blood pressure by a slower-developing mechanism: infusion of angiotensin II at a rate too low to have a direct vasoconstrictor effect raises the blood pressure gradually and sometimes markedly. Some workers consider that this second effect is more important in renal artery stenosis than the direct vasoconstrictor effect of angiotensin II.

A particularly interesting form of secondary hyperaldosteronism is found in the rare condition of renin-secreting renal tumour, which occurs mainly in young patients.

In normal pregnancy, plasma levels of renin, renin-substrate, angiotensin II and aldosterone are all increased. This is an instance of physiological secondary hyperaldosteronism.

Other patterns of variation in circulating renin and aldosterone are readily predictable. Sodium loading, or the administration of sodium-retaining substances such as DOC, fluorocortisone, carbenoxolone or liquorice, depress circulating renin and aldosterone.

Primary hyperaldosteronism. An adrenocortical adenoma secreting an excess of aldosterone will lead to sodium retention, and thus to the combination of

renin suppression with elevated aldosterone. This is known as '*primary*' hyperaldosteronism (p. 1041). The same combination will also be seen in any situation where excess aldosterone secretion is stimulated by mechanisms other than the renin-angiotensin system.

Hypoaldosteronism. In Addison's disease, the sodium deficiency stimulates marked secretion of renin, but aldosterone production remains deficient despite this stimulus, because the diseased adrenal cortex is unable to respond appropriately.

Primary renin deficiency, found mainly in elderly patients, is accompanied by selective aldosterone deficiency, cortisol secretion being normal.

Direct effects of renin and angiotensin on the kidney. The renal actions of angiotensin, although undoubted, are difficult to study in isolation from the pressor and aldosterone-stimulating effects. It has been suggested that phylogenetically, renin and angiotensin may have appeared initially as components of a purely intrarenal sodium-conserving system, and that the peripheral pressor and aldosterone-stimulating actions evolved as subsequent refinements and modifications, perhaps made necessary by a terrestrial, as opposed to an aquatic or semi-aquatic, habitat.

As mentioned earlier, a small proportion of angiotensin I can be converted to angiotensin II within the kidney. The major site of such conversion is, however, the lungs, and it has been suggested that this is an adaptation preventing the accumulation of dangerously high levels of angiotensin II within the kidney. When renal blood flow is impaired, the direct renal effect of angiotensin II may be initially beneficial, preserving glomerular filtration rate, possibly by a tonic action on efferent glomerular arterioles. However, with further elevation of angiotensin II, this beneficial effect is lost, and renal failure with tubular necrosis ensues.

Many years ago Goormaghtigh suggested that a renal effect of renin might be responsible for the reduced renal blood flow and oliguria of acute renal failure (see p. 850) and that acute renal failure might be the pathological extreme of the process outlined above. This is now supported by the demonstration that very large doses of angiotensin II can produce acute renal failure with tubular necrosis in experimental animals. A wide variety of stimuli causing increases in renin secretion predispose to acute renal failure with or without renal tubular necrosis. These include cardiac failure, sodium depletion, pregnancy, haemorrhage, Addison's disease and renal artery occlusion. Conversely, sodium loading and renal denervation reduce renin levels and are thought to protect against acute renal failure.

Shock

Definition and nature of shock. Shock is the name given to the complex series of changes which result from an acute fall in cardiac output.

These changes include regulatory mechanisms which are beneficial in that they tend to maintain the circulation, particularly to those organs with the most vital and urgent perfusion requirements—the heart and the central nervous system. Unless the circulating blood volume can be restored without undue delay, the diminished blood flow through most of the tissues results in widespread impairment of cell functions, and this, together with the compensatory circulatory changes, is responsible for the clinical features of shock.

A clear distinction should be made between shock and the *fainting* or *vaso-vagal* attack. The latter is immediate, can result from all grades of injury, from severe pain, or from psychogenic stimuli such as a fright or witnessing an accident or surgical operation. Fainting is characterised by pallor, sweating, weakness, a slow pulse, marked fall in blood pressure and loss of consciousness; vomiting and convulsions may also occur. These changes last only a few minutes and recovery is rapid. The fainting attack is mentioned here because it used to be known as 'primary shock'. In fact, it is quite distinct from shock, and should not be confused with it.

Causes and types of shock

The three major causes of shock are:

1. **Reduction of blood volume**, which induces **hypovolaemic shock**: examples include *severe haemorrhage*, *extensive vascular exudation* as in burns, and conditions such as severe vomiting and diarrhoea, which cause *dehydration*.
2. **Acute cardiac failure (cardiogenic shock)**, due most often to myocardial infarction.
3. **Severe infections**, usually with bacteraemia or septicaemia, which induce **septic shock**.

Although all three major types have many features in common, they also differ in important ways. It is also important to emphasise that the longer shock persists, the more complicated it becomes, and in advanced shock all three factors—hypovolaemia, cardiac insufficiency and

bacterial infection—are often combined. It is convenient to give first an account of hypovolaemic shock, and then to describe the special features of the other types.

The nature and features of *anaphylactic shock* and of the shock-like state of *acute immune-complex disease* are described on pp. 146 and 155 respectively.

Hypovolaemic shock

This results most commonly from **acute severe haemorrhage**, due to trauma, to involvement of blood vessels in disease processes, or to a haemorrhagic disorder. Another important cause is **severe burning**, in which hypovolaemia results from inflammatory exudation of plasma fluid from the damaged small blood vessels in the vicinity of extensive burns. Thirdly, hypovolaemic shock can develop in **severe acute dehydration**, for example in association with a gastric or intestinal fistula, gastroenteritis or cholera.

Clinical features

The shocked patient is often restless and confused, has a pale, cold, sweaty skin, often with peripheral cyanosis, a rapid weak pulse, a low blood pressure, increased rate and depth of respiration, and may become drowsy and confused and finally comatose.

Haemorrhagic and traumatic shock

A normal healthy adult can lose 500 ml of blood, i.e. about 10 per cent of the blood volume, without any significant disability; the blood volume is almost restored within a few hours, although replacement of plasma proteins takes a day or two, and restoration of red cells takes much longer. Loss of 25 per cent of the blood (about 1250 ml) results in significant hypovolaemia over the next 36 hours, while a rapid loss of about half the blood volume so reduces the circulation that death is likely unless the blood volume is restored therapeutically.

Early changes. Acute hypovolaemia results in a reduced central (systemic) venous pressure

and so a diminished flow of blood into the right atrium. The stroke volume is thus lowered and the cardiac output and arterial blood pressure fall. These haemodynamic changes trigger off peripheral and central baro-receptors with consequent sympathico-adrenal stimulation, and there is a huge increase in the levels of catecholamines in the plasma, sometimes by over 200 times. As a result of impaired renal perfusion, there is also intense secretion of renin and so a great increase of angiotensin II in the plasma (pp. 258-9).

The combined effects of these massive amounts of vasoactive agents result in an increase in the tone of the systemic veins, so that in spite of their reduced content of blood, central venous pressure and right atrial filling are partially restored, the heart rate increases, and cardiac output tends to rise towards normal. The high levels of catecholamines and angiotensin also cause constriction of the arterioles and venules in the skin, splanchnic area, and indeed most of the tissues of the body, so that peripheral resistance is increased, and even without treatment *the blood pressure may be partially or fully restored, although tissue perfusion is low*. The heart and central nervous system do not suffer to the same extent as the other tissues because they can autoregulate their own perfusion: their small blood vessels do not contract in response to noradrenaline, etc., but have an inherent property of relaxing when the blood pressure falls and contracting when it rises. In consequence of this autoregulatory mechanism, *cerebral and coronary blood flow are maintained close to normal levels at blood pressures down to 50 mm Hg*. At this pressure, arteriolar relaxation is maximal and perfusion rapidly falls off at lower pressures.

This, then, is the haemodynamic status in early shock. Compensating changes have tended to keep up the cardiac output and blood pressure, and the brain and heart are preferentially supplied with blood at the expense of diminished perfusion of the other tissues. If less than 25 per cent of the blood has been lost, and if there are no serious complicating factors (see below), the blood volume will rise naturally: vasoconstriction of the arterioles is greater than in the venules, so that the pressure in the capillaries is low and extravascular fluid passes into them (p. 49), and the high levels of angiotensin II stimulate adrenal secretion of aldosterone,

which promotes retention of salt and water. The circulation is nevertheless precarious, and further bleeding, major surgery to deal with the causal injury or bleeding vessel, severe pain, or the development of infection, will all tend to increase the circulatory deficit. *It is therefore important, in all save the mildest cases, to restore the blood volume by intravenous administration of fluid*. The nature of the fluid is not so important as the avoidance of delay: buffered saline or macromolecular solutions (plasma, dextran, etc.) are both effective initially, but macromolecular solutions have the advantage of maintaining the osmotic pressure of the plasma, thus tending to hold fluid in circulation, and are usually used for losses of around 25 per cent or more of the blood. It is also important to maintain the haematocrit at around 30 per cent in order to minimise tissue hypoxia, and matched blood (or in an urgent situation Group O Rh negative blood) are normally administered if haemorrhage has exceeded 25 per cent of the blood volume. Some estimate of the volume of fluids required can be made from the amount of blood lost, the clinical state, and the severity and nature of injury, but account must also be taken of internal haemorrhage, e.g. into the gastro-intestinal tract or around a fracture. The haemoglobin and haematocrit levels are not reliable guides to the degree of hypovolaemia during the first 36 hours. In the absence of cardiac insufficiency, a low blood pressure is an indication of hypovolaemia in early shock, but because of the compensatory mechanisms described above, it may be normal or nearly so in patients with serious hypovolaemia. A low central venous pressure is often, although not always, a useful indication of hypovolaemia, and if possible this should be monitored in all except mild cases of shock.

Although the peripheral vasoconstriction of shock serves a compensatory function, it is also harmful by reducing general tissue perfusion and it may, by increasing peripheral resistance, induce heart failure (see below). In some cases, the blood pressure may rise above normal, and the vasoconstriction may persist in spite of restoration of the blood volume. Drugs which promote vasodilatation (e.g. thymoxamine, sodium nitroprusside) are therefore sometimes beneficial, but only when steps have been taken to restore the blood volume: in the hypovol-

aemic patient they are liable to cause further circulatory collapse.

The changes of advanced shock. If shock persists, the widespread arteriolar constriction gradually passes off, but venular constriction is more persistent and capillary pressure rises with consequent loss of fluid into the extravascular space and further fall in blood volume. At this late stage of shock, the capillaries are congested with slowly-flowing blood, and cyanosis may be apparent. The general reduction in blood supply to the tissues is aggravated by a number of complex factors brought about by changes in the blood itself and by the injury to vascular endothelium and tissue cells resulting from perfusion failure. Some of the changes are as follows:

(a) *Viscosity of the blood* is increased by the haemoconcentration resulting from loss of capillary fluid. This leads to sludging of the red cells and rouleaux formation (p. 47) and these effects are increased by the rise in plasma fibrinogen which follows haemorrhage.

(b) *Release of thromboplastin* (Factor III) from hypoxic endothelium and tissue cells results in the production of thrombin (p. 234), which promotes aggregation of platelets and occasionally intravascular formation of fibrin. Aggregated platelets release adenosine diphosphate and thromboxane A_2 which cause further platelet aggregation.

(c) *Neutrophil polymorphs* adhere to the injured vascular endothelium of small vessels.

(d) *Hypoxic injury* results in release of lysosomal enzymes and secretory products into the blood. Proteolytic enzymes, e.g. trypsin from the pancreas, may activate the kinin system and thus further embarrass the circulation by causing vasodilatation and increased permeability. Production of the prostaglandins may also be increased: those of the E group have a kinin-like effect, while the F group may increase the resistance to pulmonary blood flow.

Metabolic disturbances. The hypoxia of shock interferes profoundly with cell metabolism. It prevents the entrance of pyruvic acid into the citric acid cycle and in consequence lactic acid accumulates and glucose passes out of the hypoxic cells, leading to insulin-resistant

hyperglycaemia and increased glycogenolysis. These metabolic disturbances together with high levels of catecholamines, result in a rise of fatty acids and amino acids in the plasma. Impaired carbohydrate metabolism results in a fall in production of adenosine triphosphate and so energy is not available for many cell functions, including the sodium pump: *potassium leaves the cells and sodium and water enter and cause swelling: these effects, sometimes termed the 'sick cell syndrome' (p. 19) may, by lowering the level of blood sodium, lead to inappropriate administration of salt.*

Metabolic acidosis, with rise in blood lactic acid, contributes to the hyperventilation of shock.

Organ function in shock. While all the organs are affected in shock, respiratory and cardiac failure are commonly of life-threatening importance. Quite apart from cardiogenic shock (see below), **acute heart failure**, first of the left and then of both ventricles, may develop in severe hypovolaemic or septic shock, and is particularly common in older patients with pre-existing coronary artery disease. The increased load on the heart resulting from peripheral vasoconstriction and its treatment with vasodilator drugs has been considered above. A factor which reduces myocardial contractility (*myocardial depressant factor*) has been detected in the plasma of shocked patients who subsequently died of cardiac failure: it is believed to be released from the pancreas. The impaired blood flow of severe shock, together with activation of the clotting mechanism, predispose to coronary thrombosis in patients with coronary artery disease. If operation is necessary, anaesthetic drugs may also impair cardiac function. Monitoring of the cardiac filling* and systemic arterial pressures, and particularly of changes in them during intravascular administration of fluid, helps to distinguish between hypovolaemia and cardiac insufficiency in shock. In some cases, drugs such as dopamine or digitalis, which increase myocardial contractility, are beneficial.

Disturbance of gas exchange in the lungs is another important complication of shock, and can be assessed by comparing the mixed venous and arterial oxygen tensions. Improvement

* The central venous pressure is used as a measure of the right heart filling pressure and the pulmonary artery 'wedge' pressure as an indication of the left heart filling pressure.

usually follows restoration of the blood volume, together with intermittent positive-pressure ventilation if necessary, but in some cases pulmonary function continues to deteriorate due to a combination of causes—pulmonary oedema, alveolar collapse, intravascular fibrin formation, embolism, infection, etc., known collectively as **shock lung** (p. 454), and death is then likely to result largely from the additional burden of respiratory failure.

Perfusion of the **kidneys** in shock is directly proportional to the blood pressure. Production of urine ceases at about 50 mm Hg and if the pressure remains low for some hours, focal hypoxic injury to the tubular epithelium may be associated with acute renal failure which persists for days or weeks after recovery from shock (p. 849). Renal damage is particularly common in shock associated with crush injury, childbirth, incompatible blood transfusion or severe infection.

Because of its autoregulatory mechanism, blood flow to the **brain** is relatively well-maintained unless the blood pressure falls below 50 mm Hg. Even a brief period of more severe hypotension can cause severe ischaemic brain damage (p. 247). Ischaemic centrilobular necrosis of liver cells may also occur, although liver failure is seldom prominent.

Other causes of hypovolaemic shock

Burns. In burning or scalding, necrosis of the more superficial tissues is accompanied by a lesser degree of injury to the underlying tissues, the reaction to which is acute inflammation. The small vessels dilate and their permeability increases, so that there is exudation of protein-rich fluid. When the area involved is extensive (10 per cent or more of the skin surface), the loss of fluid is severe enough to induce hypovolaemia and shock. The changes are similar to those following haemorrhage but hypovolaemia develops more slowly, haemoconcentration is more pronounced, with its attendant sludging and rouleaux formation, there is usually a marked leukocytosis, and the state of shock may recur or increase on the second or third days, possibly as a result of infection or absorption of breakdown products from the necrotic tissue. The principles of treatment of burn shock are the same as for haemorrhagic

shock, but the loss is of plasma rather than whole blood, so that plasma transfusions are used initially. Some destruction of red cells does however, occur in the burned area, and in very extensive burning severe anaemia may develop, necessitating blood transfusion. Another important complication of burning is bacterial infection, the dead tissue providing a good culture medium from which bacteria commonly invade the underlying tissue and bloodstream. Streptococci and staphylococci were formerly the most important invaders, but with antibiotic therapy Gram-ve bacilli, and especially *Pseudomonas aeruginosa*, now predominate. The features of septic shock commonly supervene, and sepsis is now the major cause of death from burns.

Dehydration, if severe, causes hypovolaemic shock, although the blood volume is reduced relatively less than the extravascular fluid, and the effects of cellular dehydration are regarded as the usual cause of death (p. 252).

Cardiogenic shock

Acute lesions of the heart may severely reduce cardiac output and the subsequent haemodynamic and other changes are similar to those in hypovolaemic shock. The cardiac filling pressures are, however, raised, and although the clinical features—pallor, weak rapid pulse, sweating, etc.—are the same as in hypovolaemia, intravenous administration of fluids, which is beneficial in selected cases, must proceed with caution.

The commonest cause is myocardial infarction, and although only a small proportion of patients with this condition develop the full picture of shock, the mortality from this complication is very high even if facilities are available for intensive therapy. Other conditions which can cause cardiogenic shock include rupture of a valve cusp, major arrhythmias, and cardiac tamponade due to haemopericardium (resulting from direct trauma or as a complication of a ruptured myocardial infarct). Although not strictly cardiogenic shock, acute obstruction to blood flow by pulmonary emboli can result in a similar condition. As already indicated, cardiac insufficiency can develop as a complication of haemorrhagic and other types of shock.

Septic shock

Some patients with septicaemia or extensive localised infections, such as general peritonitis, pass into a state of shock which resembles that of hypovolaemia, but is often more prolonged, with a higher incidence of serious complications, and an overall mortality exceeding 50 per cent.

Septic shock is a common complication of infected burns, and of surgical operations, manipulations or instrumentation on the urogenital, gastro-intestinal and biliary tracts. It occurs also in patients with immunodeficiency states, such as leukaemia and lymphomas, and as a complication of cytotoxic drugs or immunosuppressive therapy. Many such patients are dehydrated, and this is an important predisposing factor.

In a patient with known sepsis, high fever and symptoms and signs of shock, the diagnosis is not difficult, but in many patients, especially the elderly and following surgical procedures, septic shock develops insidiously, often without fever, and there may be an initial 'hyperdynamic' stage in which cardiac output is increased, the blood pressure reduced, peripheral resistance low, and the skin warm: these features may be due to bacteria-mediated release of kinins and other vaso-active agents. More often circulatory changes are similar from the onset to those of hypovolaemia, with pallor, sweating, cold extremities, increased peripheral resistance and reduced cardiac output, etc. Septic shock is particularly difficult to reverse and, as with other forms of shock, the longer it persists the more refractory it becomes. Tissue hypoxia results in widespread derangement of cell function, and features of multi-organ failure often develop. **Respiratory failure** due to 'shock lung' is often combined with **cardiac failure** and arrhythmias. **Acute renal failure** is also very common, although (as in hypovolaemic shock) many of its serious effects develop after recovery from shock.

Disseminated intravascular coagulation (p. 556) is more prone to develop in septic than in hypovolaemic shock. It interferes further with organ perfusion and may greatly aggravate pulmonary failure, or may, by consumption of clotting factors and activation of the plasmin system, progress to a bleeding state with wide-

spread haemorrhage, e.g. from the gastrointestinal mucosa.

Treatment of septic shock is based on elimination of the causal infection, restoration of the circulation, and correction of hypoxaemic metabolic acidosis and electrolyte imbalance. Antibiotic therapy cannot usually await the results of bacteriological culture of the blood etc. and tests for sensitivity, although it may need to be subsequently modified according to the bacteriological findings. The choice and rate of administration of intravascular fluids will depend on the results of monitoring procedures and haematological and biochemical tests. These include monitoring the cardiac filling and systemic arterial pressures, and frequent assay of the blood gases and pH, plasma electrolytes and osmolality, platelet counts and haematocrit. It may also be necessary to assay the status of the coagulation and plasmin systems. Intermittent positive-pressure ventilation, instituted at an early stage, has been shown to decrease the risk of the development of pulmonary failure, and drugs which increase cardiac output and tissue perfusion may be of value. The prognosis depends very much on the availability of experienced staff and facilities.

Aetiology. The microbial factors responsible for septic shock are by no means fully elucidated. With widespread use of antibiotics, the aerobic Gram-ve bacilli have replaced the pyogenic cocci as the major cause of septicaemia and bacteraemia. The organisms most commonly responsible include *Esch. coli*, *Proteus*, *Klebsiella* and in cases of burns *Pseudomonas aeruginosa*. About 50 per cent of patients with blood infection by these bacteria develop septic shock. Bacteroides (the anaerobic non-sporing bacilli which constitute over 99 per cent of the faecal flora) have been recognised quite recently as an important cause of blood infection, and about 30 per cent of cases are complicated by shock. All these Gram-ve bacteria release endotoxins when they die, and there is a widespread belief that endotoxin is a major cause of the manifestations of septic shock. Animals injected with endotoxin present many of the features of septic shock, including disseminated intravascular coagulation (the so-called *Schwartzman reaction*), and endotoxin has been detected in the blood of shocked patients with Gram-ve septicaemia by means of the limulus test (in which endotoxin is detected by its property of clotting a lysate of the blood amoebocytes of *Limulus polyphemus*, the horse-shoe

crab). The limulus test is, however, time-consuming and, unless performed by an expert, can be misleading.

Endotoxin may induce the features of septic shock by a combination of its known effects, including activation of complement by the alternative pathway (p. 143), by promoting intravascular clotting, by its cytotoxic effects on neutrophil polymorphs, or possibly by promoting the production of kinins, prostaglandins, etc.

Severe shock sometimes develops in patients with fungal or acute virus infections, and in some instances immune-complex formation may be involved.

Other causes of shock

Some cases of shock do not fall into any of the three major types described above. For example, escape of gastric or duodenal juice into the peritoneal cavity, via a perforated peptic ulcer, causes severe shock, and so does acute haemorrhagic pancreatitis (which is non-bacterial), and the drinking of many poisons. In these conditions, it is likely that shock is chemically-induced, but there may also be severe pain which, without doubt, aggravates shock.

Severe shock results from transfusion of strongly incompatible blood to which the recipient has iso-antibodies; also in acute circulating immune complex disease (p. 155) and acute generalised anaphylaxis (p. 146).

Morphological changes in shock

The morphological changes in patients dying from shock are often inconspicuous. In spite of the fundamental disturbances of cell function, the parenchymal cells in general usually show only swelling and sometimes fatty change. In addition to the causal changes—injury, haemorrhage, coronary thrombosis, septicaemia, etc.—there may be various pulmonary changes, including oedema, congestion, hyaline membrane formation, collapse and bronchopneumonia (p. 454). The kidneys usually show the pallor and cortical swelling of acute tubular injury (p. 848) and there may be centrilobular hepatic necrosis. If the patient has survived sufficiently long for its recognition, there may be

acute ischaemic necrosis in the 'boundary' zones of the brain (pp. 745–6). Features of disseminated intravascular coagulation include widespread haemorrhages, and microscopy may reveal fibrin thrombi in the small vessels, especially in the lungs and kidneys (Fig. 9.35). The adrenals show the lipid depletion of the 'stress' reaction (p. 1039), but occasionally there is a combination of haemorrhage and necrosis, particularly in septic shock associated with meningococcal septicaemia.



Fig. 9.35 Fibrin thrombus in a small pulmonary vessel in a case of disseminated intravascular coagulation. $\times 200$.

Metabolism after injury

The metabolic disturbances associated with shock (p. 262) include incomplete carbohydrate catabolism, metabolic acidosis, disturbed protein and fat metabolism and a rise in the blood levels of glucose, amino acids and fatty acids. These changes were demonstrated experimentally by Cuthbertson, who termed them the '*ebb phase*'. Energy production is consequently depressed, and there is a general disorder of cellular metabolic processes. These changes are liable to develop in the first days following a severe injury, burn or surgical operation. Following the period of shock (or 2–3 days after such injury when shock has been prevented), the metabolism changes to a '*flow phase*' in which there is increased energy (and heat) production, due largely to breakdown of depot fat and protein, with consequent loss of weight and a negative nitrogen balance. During this period, which may persist for days to months

depending on the severity of the injury, carbohydrate catabolism is complete and there is no metabolic acidosis unless carbohydrate intake is low. The mechanism of the increased

metabolic activity, which resembles that in fever, is uncertain, but weight loss can be minimised by a high calorie, high protein diet, and a warm environment.

Blood Groups and Blood Transfusion

Before administering a blood transfusion, it is essential to make sure that the donor's red cells are compatible to the patient, and in particular that the patient's plasma does not contain iso-antibodies reactive with surface antigens on the donor's red cells. The ABO blood group system is of outstanding importance, for iso-antibodies are normally present in the plasma (see Table 9.2) and their reaction with incompatible transfused red cells usually causes a severe haemolytic reaction with fever, shock, often acute renal failure and sometimes death. The rhesus (Rh) blood group system comes next in importance. Unlike the ABO system, Rh iso-antibodies are not usually present in the plasma, but they sometimes develop as a result of an Rh-incompatible blood transfusion or pregnancy.

For blood transfusion, the patient's ABO and Rh types should be determined and blood of the same type should be selected for transfusion. In addition, it is necessary to perform a *compatibility* or *cross-matching* test in which the donor's red cells are incubated in the recipient's serum at 37°C, and the cells are then examined for agglutination and also by the anti-globulin test (p. 111) to detect non-agglutinating (IgG) antibodies in the recipient's serum. The purpose of this procedure is to detect (1) technical or clerical errors in grouping and in collection and storage of the donor's blood,

and (2) the presence of unusual iso-antibodies in the recipient's plasma.

As a life-saving measure, it may be appropriate to transfuse Group O, Rh-ve blood to a patient of unknown group, but it is usually preferable to administer plasma while grouping procedures are being performed, and in any case a direct cross-matching procedure should always be performed.

Incompatibility can also arise when a donor's blood contains high titre antibodies reactive with the patient's red cells. Transfusion reactions from this cause are not usually severe because the donor's plasma (and thus the antibody) is diluted *in vivo* by the recipient's plasma. Screening tests for high titre ABO antibodies are not performed by most transfusion centres, and if donor and recipient are of the same ABO and Rh type the danger is largely excluded.

The ABO groups

Individuals can be classified into four groups by the presence or absence of A and B antigens on their red cells and of anti-A and anti-B antibodies in their plasma (or serum). Table 9.2 shows the features of the four groups.

Standardised anti-A and anti-B serum (from selected subjects of group B and A respectively)

Table 9.2 The four ABO blood groups

Blood group	Red cell antigens	Iso-antibodies in serum	Can accept blood of group	Can donate to patients of group	Incidence in Britain* (%)
AB	AB	nil	all groups	AB	3
A	A	anti-B	A, O	A, AB	42
B	B	anti-A	B, O	B, AB	8
O	O	anti-A + anti-B	O	all groups	47

*The frequencies of the four groups vary greatly in different ethnic groups.

are used to determine the group to which any individual belongs. If the red cells are agglutinated by both sera the blood belongs to group AB, if agglutinated by group B serum alone the blood belongs to group A, if by group A serum alone to group B, and if by neither serum, it belongs to group O. Since the serum of group AB does not agglutinate the red cells of any of the groups an individual of group AB can receive the red cells of any other group and is thus a 'universal recipient'. The cells of an individual of group O are not agglutinated by the serum of any group; the red cells can be transfused into an individual of any group and such persons are known as 'universal donors'.

The three blood group substances A, B and O are determined by allelic genes, one from each parent, so that there are six genotypes (AA, BB, AB, AO, BO and OO). O substance, however, is for practical purposes non-antigenic, and accordingly grouping is based on the presence or absence of A and B, giving the four phenotypes, which were first detected by Landsteiner. The iso-antibodies are mainly of IgM class, and develop after birth, apparently as a result of exposure to bacterial and other substances antigenically similar to A and B. Group A (or B) individuals are immunologically tolerant to A (or B) and so do not develop the corresponding antibodies.

The Rhesus (Rh) groups

The Rh blood group system was discovered by Landsteiner and Wiener (1940), who were interested in the antigens of human and animal red cells, and noted that guinea-pig or rabbit antisera to the red cells of *Macacus rhesus* monkeys agglutinated the red cells of 84 per cent of white Americans; accordingly, these 84 per cent were called Rh-positive, and the 16 per cent of non-reactors, Rh-negative. The human Rh system is, however, more complex, and further elucidation has come from the use of iso-antibodies which, unlike the ABO antibodies, are not routinely present in human serum, but develop in about 50 per cent of Rh-ve subjects transfused with Rh+ve blood, and in about 5 per cent of Rh-ve women as a result of an Rh+ve pregnancy (p. 151). Once Rh antibodies have developed, a subsequent Rh+ve blood transfusion is likely to cause an acute reaction with immune destruction of the trans-

fused red cells by a cytotoxic antibody (type II) reaction, while an Rh+ve fetus is liable to suffer from haemolytic disease of the newborn (p. 527).

Rhesus iso-antibodies may be of either IgM or IgG class: their demonstration requires incubation with appropriate red cells at 37 °C. Since IgG antibodies sensitise the red cells without agglutinating them, their detection is usually effected by use of the antiglobulin reaction (p. 111) and by methods which render red cells agglutinable by Rh antibodies, e.g. treating the red cells with papain or suspending them in a concentrated albumin solution.

Rh sub-groups. In simple terms, Rh blood group antigens are determined by three pairs of allelic genes, one of which codes for antigens C and c, one for D and d and the third for E and e. As the genes are closely linked, they are transmitted as haplotypic 'sets' which may be expressed as CDe, cde, etc., or by a set of symbols (R_1 , r , etc.). The frequency of the eight possible haplotypes varies considerably in different peoples: their frequency in this country, together with the alternative symbols are shown in table 9.3

Table 9.3 The major Rh haplotypes and their frequency in Britain.

Haplotype	Abbreviation	Frequency*†
CDe	R_1	0.420
cde	r	0.389
cDE	R_2	0.141
cDe	R_0	0.026
cdE	r''	0.012
Cde	r'	0.010
CDE	R_z	very rare
CdE	r_y	very rare

* The frequency varies greatly in different ethnic groups.

† The frequency of any given *genotype* is obtained by multiplying together the frequencies of the two haplotypes as given here: thus CDe/cde occurs in $0.420 \times 0.389 = \text{approx. } 16 \text{ per cent}$ of the population of Britain, and cde/cde in $0.389^2 = 15 \text{ per cent}$. Reversing the calculation gives an estimate of gene frequency for known genotype frequencies.

Each individual inherits one of these sets from each parent, and his red cells may thus have from 3 to 6 different Rh antigens. In practice, iso-immunisation develops mainly when an individual of genotype cde/cde receives red cells which are D-positive. Accordingly, indivi-

duals whose red cells possess D are termed *Rh-positive* and those without D (i.e. with dd) are termed *Rh-negative*. In Caucasian stock, about 15 per cent of individuals are cde/cde (hence the frequency of cde as calculated in the table is $\sqrt{\frac{15}{100}} = 0.389$). Rh-ve individuals with other genotypes are comparatively rare.

There is also a significant risk of the development of anti-C when C+ cells are transfused to an Rh-ve individual, and it is common practice to use a mixture of anti-C and anti-D for Rh grouping. Iso-immunisation may also result when c+ or e+ blood is transfused to CC or EE individuals respectively, but d+ blood does not iso-immunise DD individuals.

The Rh system is, in fact, much more complex than has been suggested above: a fourth

antigen, G, is closely associated with C and D, and further antigens, determined by variants of the common allelic genes or joint products of the genes, also occur.

Other blood group systems

In addition to the ABO and Rh groups many other blood group systems are known, but like the Rh group natural iso-antibodies are absent, and the sera that detect these groups are obtained mostly from persons immunised by transfusion or by pregnancy. These groups only rarely bring about iso-immunisation. Nevertheless the greatly increased use of blood transfusion necessitates their consideration and identification when a cross-matching test reveals an unexpected antibody.

References

-
- Masotti, G., Poggesi, L., Galanti, G., Abbate, R. and Neri Serneri, G. G. (1979). Differential inhibition of prostacyclin production and platelet aggregation by aspirin. *Lancet* ii, 1213-6.
- Moncada, S. and Vane, J. R. (1979). Arachidonic acid metabolites and the interactions between platelets and blood-vessel walls. *New England Journal of Medicine* 300, 1142-7.
- Sevitt, S. (1973). The mechanisms of canalisation of deep vein thrombosis. *Journal of Pathology* 110, 153-65.
- Sevitt, S. (1973). The vascularisation of deep vein thrombi and their fibrous residue: a post-mortem angiographic study. *Journal of Pathology* 111, 1-11.
- Singer, F. R., Schiller, A. L., Pyle, E. B. and Krane, S. M. (1978). Paget's disease of bone. In *Metabolic Bone Disease*, Vol. 2, pp. 548-9. Ed. by L. V. Avioli and S. M. Krane. Academic Press, New York, London and San Francisco.

Further Reading

-
- Mollison, P. L. (1978). *Blood Transfusion in Clinical Medicine*, 6th edn. pp. 896. Blackwell Scientific, Oxford.
- Race, R. R. and Sanger, Ruth (1975). *Blood Groups in Man*, 6th edn. pp. 682. Blackwell Scientific, Oxford.
- Sevitt, S. (1974). *Reactions to Injury and Burns and their Clinical Importance*, pp. 256. Heinemann Medical, London.
- Thomas, D. (Ed.) (1977). Haemostasis. *British Medical Bulletin* 33, 118-288. (Reviews by leading workers.)
- Thomas, D. (Ed.) (1978). Thrombosis. *British Medical Bulletin* 34, 101-207. (Reviews by leading workers.)

Miscellaneous Tissue Degenerations and Deposits

The degenerative changes which result from cellular injury, and the intracellular accumulation of lipids and glycogen resulting from certain disorders of metabolism, have been dealt with in Chapter 1. In the present chapter we describe a group of changes which, although heterogeneous, consist of either the accumula-

tion in the tissues of various substances—amyloid material, mucus, pigmented compounds, calcium deposits and urates—or tissue degenerations which usually affect the stroma of supporting tissues, and are recognised by their microscopic appearances but are ill-defined chemically.

Amyloidosis

Amyloid is a predominantly extracellular fibrillar material composed essentially of protein: it is deposited in various tissues in a number of diseases. Extensive deposits are visible by naked eye, causing enlargement of the involved organ and giving it a waxy appearance (Fig. 10.1). In

sections stained by haematoxylin and eosin amyloid is seen as a homogeneous, pink, refractile material. Electron microscopy shows it to consist of filaments of 7.5 nm diameter (Fig. 10.2), which can be dissociated into several protofibrils and are often twisted together in pairs.



Fig. 10.1 Amyloidosis of liver. The amyloid material renders the organ firm, and gives it a dark, homogeneous appearance. $\times 1$.

Methods of demonstrating amyloid

The wide variety of methods currently used to demonstrate the presence of amyloid is an indication of their lack of specificity. Large deposits of amyloid are usually demonstrable by all the methods. If the amount of amyloid present is small, however, the results with each method vary from case to case and identification is correspondingly difficult; for this reason it is usual to use several methods, the best known of which are as follows.

(1) **Lugol's iodine.** Amyloid has a strong affinity for iodine (hence its name) and this forms the basis for a useful macroscopic test (Fig. 22.51, p. 842). When Lugol's iodine solution is poured over tissue, the amyloid is stained deep brown in contrast to the normal tissue which is only lightly stained. Congested tissues should first be rinsed free of excess blood as this obscures the test.

(2) **Congo red.** This stain may be used on gross specimens and sections for microscopy. Formerly the



Fig. 10.2 Electron micrograph showing renal amyloidosis. The field shows glomerular capillary basement membrane on the outer side of which (above) the foot processes of the epithelium have fused to form a continuous layer. The inner part of the basement membrane is irregularly permeated by amyloid which also occupies the sub-endothelial space (lower half of the field) and is seen as fine filaments. $\times 39\,000$.

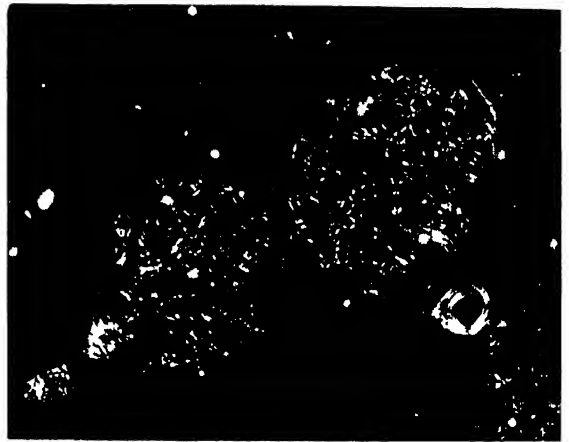
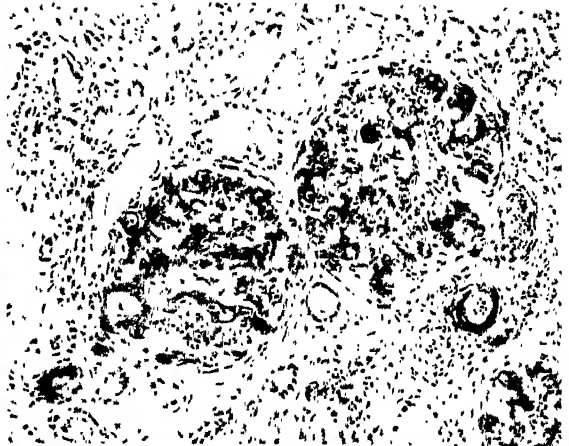


Fig. 10.3 The kidney in amyloidosis, stained by Congo red. The glomerular capillaries and the arterioles are affected. *Above*, viewed by ordinary light: the amyloid is seen as homogeneous material. *Below*, viewed by crossed polarising films, showing birefringence of the amyloid. $\times 105$.

rate of disappearance from the serum of an intravenously injected solution of congo red was used as a clinical test for amyloidosis but owing to severe reactions and lack of accuracy it was abandoned. In polarised light amyloid stained by congo red shows a green birefringence, and this is the most reliable technique (apart from electron microscopy) for demonstrating amyloid (Fig. 10.3).

(3) Rosaniline dyes. These include gentian violet, methyl violet and crystal violet; they stain amyloid reddish while other tissue elements appear purple. This phenomenon of a dye reacting with a tissue constituent and undergoing a colour change is called *metachromasia*: it is believed that in this instance it is due to selective binding of impurities in the dyes by amyloid fibrils.

(4) Fluorescent dyes. Thioflavine-T binds to amyloid and its presence is demonstrated by fluorescence microscopy. The reaction is not, however, entirely specific for amyloid.

The deposition and effects of amyloid

Amyloid is deposited extracellularly, and first appears in the walls of small vessels, both arterial and venous, in relation to the basement membrane of capillaries and vascular sinusoids, and also of epithelial structures, e.g. the renal tubular basement membrane.

When present in small amounts, amyloid has little effect on the organs, with the exception of the kidneys in which quite early glomerular deposition may result in proteinuria (see below). Involved small vessels tend to be susceptible to trauma and to bleed readily, giving rise to petechial haemorrhages.

In greater amounts, amyloid causes enlargement of the organs and imparts to them a firm

rubbery consistency and a slightly translucent, waxy appearance. The major features of involvement of the individual organs are as follows.

Amyloid produces effects by pressure on adjacent cells and by interfering with the normal transfer of water and solutes across the walls of affected small blood vessels.

The liver is firm and elastic and may be palpable during life. Amyloid appears to be deposited first in the space of Disse (between the sinusoidal endothelium and the hepatocytes) and as it increases forms a continuous network between the sinusoids and columns of liver cells. The change usually begins in the sinusoids of the intermediate zones of the lobules: it may become very extensive and produce marked atrophy of the liver cells (Fig. 10.4). Even at an advanced stage, liver function is not usually severely impaired.



Fig. 10.4 Amyloidosis of the liver. The pale homogeneous amyloid substance extends around the walls of the sinusoids, enclosing the columns of liver cells, which are undergoing atrophy. The zone around the central vein (*top right*) is least affected. $\times 115$.

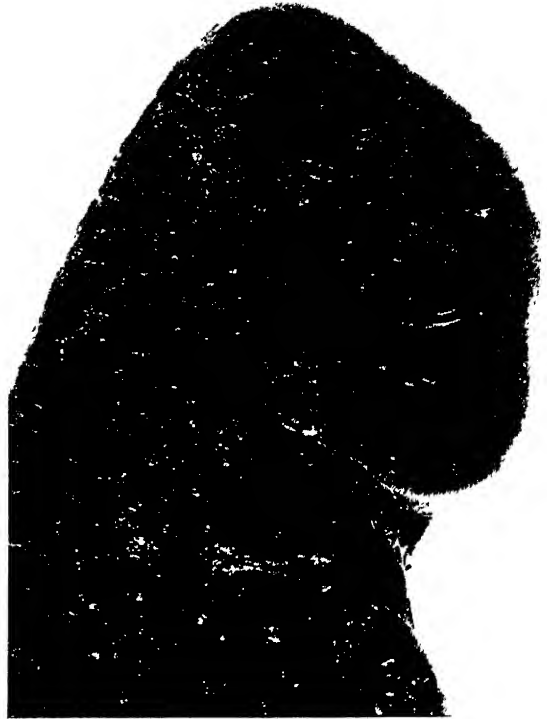


Fig. 10.5 Amyloid spleen of 'sago' type, i.e. affecting the Malpighian bodies.

The spleen shows two distinct patterns of involvement. In one the Malpighian bodies are changed to translucent globules by amyloid deposition in their reticulum (Fig. 10.5); hence the term 'sago spleen'. In this form splenomegaly is not marked. In the diffuse form, the change affects reticulum of the red pulp, walls of venous sinuses, and many of the small arteries; the spleen may be palpable and weigh up to 1 kg. The explanation of the two distributions is not known.

Amyloidosis of the kidneys is particularly important because of its effect on renal function. Deposition occurs upon the basement membranes of the tubules and in the walls of arterioles (Fig. 10.3) and venules, but the most important site is in relation to the glomerular capillary basement membrane (Figs. 10.2, 22.49 and 22.50, p. 842): amyloid is deposited initially on the endothelial side of the basement membrane, but extends through it to accumulate also on the epithelial side. The glomerular capillaries are rendered abnormally permeable to macromolecules, with consequent heavy proteinuria and nephrotic syndrome (p. 256).

Eventually many of the glomerular capil-

laries are obliterated, the kidneys become scarred and shrunken and chronic renal failure develops.

In the **stomach and intestines**, amyloid deposits may be widespread and this leads to atrophic changes in the mucosa. Diarrhoea may result from severe involvement of the gut, but even in its absence amyloid material is often demonstrable in rectal biopsy material, providing a useful diagnostic measure. Gingival biopsy is also useful, although less often diagnostic than rectal biopsy.

Other organs. Deposits of amyloid may also be found in the adrenals, heart, thyroid, skin and lymph nodes, and indeed in almost any tissue.

In the **heart**, it is deposited subendocardially and irregularly between myocardial fibres. Even when it has caused heart failure, there may be nothing to suggest its presence on naked-eye examination, but sometimes the ventricular myocardium is thickened and firm.

Classification of amyloidosis

Amyloidosis has been classified in several ways, none of which is entirely satisfactory. This is largely because, whichever method of classification is used, there is considerable overlap in the features which distinguish the different types, and while general rules can be applied, many exceptions are encountered. It seems at present more appropriate to distinguish between the two major types of amyloid material now known to exist, and also between generalised and localised amyloidosis. Lastly, the familial forms of amyloidosis are sufficiently distinctive to place in a separate group.

The two forms of amyloid material will be discussed later. Suffice it to say now that one form may be derived from a plasma protein but its real nature is uncertain and it is best termed AUO (*amyloid of unknown origin*); the other form is derived from the light chains of immunoglobulin and is conveniently termed AIO (*amyloid of immunoglobulin origin*).

Amyloidosis due to AUO. This is the classical form of **secondary amyloidosis** and develops in diseases characterised by chronic destructive inflammatory lesions and in certain forms of cancer. It is a common complication of tuberculosis, lepromatous leprosy, syphilis,

chronic bacterial osteitis, bronchiectasis and chronic suppurative pyelonephritis. These conditions have become amenable to treatment, and are less commonly seen as causes of amyloidosis in developed countries. Amyloid of unknown origin is also seen as a complication of rheumatoid and other forms of chronic arthritis, in which it can be detected at necropsy in about 20 per cent of cases, in Hodgkin's disease (a neoplastic condition, probably of macrophages), and occasionally in chronic ulcerative conditions of the skin (e.g. bedsores) or of the intestine (e.g. Crohn's disease), systemic lupus erythematosus and carcinomas.

The amyloid material is deposited most often and in greater amount in the liver, spleen, kidneys and adrenals, but has a wide distribution and some degree of involvement of the alimentary tract and lymph nodes is common. The major cause of death is renal failure.

Amyloid of unknown origin is usually typical in its staining reactions, being readily demonstrated by the methods outlined above.

Amyloidosis due to AIO. This occurs in approximately 15 per cent of patients with multiple myeloma (a neoplasm of the bone marrow of plasma cell type), in which the amyloid is almost certainly derived from light chains secreted by the neoplastic plasma cells; it may also complicate other neoplastic conditions of plasma cell type. It also occurs as a **primary condition**, almost always in old people: in some such cases, there is evidence of a latent form of plasma cell neoplasia, but in others no predisposing cause can be found. Typically, the distribution differs from that of AUO amyloidosis, the tissues severely affected being, in order of frequency, the heart, alimentary tract (including the tongue), skin, skeletal muscles, spleen, kidneys, liver and lungs. Death commonly results from myocardial failure, often with arrhythmias. Careful microscopic examination of necropsy material reveals minor degrees of this form of amyloidosis in a small percentage of old people dying from various causes.

Amyloid of immunoglobulin origin is often atypical in its staining reactions, and when present in small amounts it may be difficult to demonstrate convincingly.

Localised amyloidosis, restricted to one organ or tissue, is relatively common in the larynx where it gives rise to small tumour-like nodules. It may be restricted to the skin, bron-

chi, lungs, heart or urinary bladder, and is a rare cause of enlargement of the thyroid. It occurs also in the pancreatic islets in many diabetics, and as a feature of certain tumours, e.g. medullary thyroid cancer, islet cell pancreatic tumours and phaeochromocytomas of the adrenal medulla.

Genetically-determined amyloidosis. Several familial forms of amyloidosis have been described. The best known are familial mediterranean fever and primary familial amyloidosis.

Familial mediterranean fever is found principally in Mediterranean Jews and Armenians and is inherited as an autosomal recessive disease. In its most typical form, recurrent fever is associated with pain in chest, abdomen, joints and skin; amyloidosis supervenes, causing death by renal involvement, but affecting also the spleen, lungs and liver. Variants of the disease are recognised in which the amyloidosis becomes apparent before the other features.

Primary familial amyloidosis. This is least rare in parts of Portugal and is inherited as an autosomal dominant. The disease presents in the 3rd and 4th decades with increasing leg weakness and loss of reflexes. Subsequently sphincteric disturbances and malabsorption from intestinal involvement lead to death within 10 years.

Many other rare syndromes have been described in individual families, each with a particular distribution of amyloid deposition.

The nature and aetiology of amyloidosis

The essential constituent of amyloid is the protein which forms the fibrils seen on electron microscopy. Investigation of AIO, including amino-acid sequencing and immunological analysis, has shown it to have a molecular weight of 5000 to 18 000 and to consist of the N-terminal parts of immunoglobulin light chains. In any particular case, the amyloid is of *either* λ or, less commonly, of κ type (p. 106) and amino-acid sequencing indicates homogeneity of the polypeptide chains, which is strong evidence that *it is produced by a clone of plasma cells*. Amyloid from different patients shows differences in its amino-acid sequences. It is significant that nearly all patients with multiple myeloma complicated by amyloid have free light chains in their plasma and urine, and that partial proteolytic digestion of light chains *in vitro* produces a breakdown product which assumes the structure of amyloid filaments.

Analysis of AUO has shown it to be homo-

geneous in each individual case, but there is disagreement on whether it too consists of a part of the immunoglobulin molecule. On balance, this appears to be unlikely, and antibodies to AUO have been reported to cross react with a trace protein present in the plasma, the concentration of which increases in old age. It may be that AUO is a breakdown product of this.

A characteristic feature of amyloid, revealed by x-ray diffraction analysis, is that the polypeptide molecules have a β -pleated configuration, which is unusual for mammalian polypeptides, and may explain its chemically inert nature, insolubility and poor antigenicity. It is of interest that, on enzymic digestion, a number of polypeptides, including insulin, glucagon and calcitonin yield fragments which can assume this configuration and adopt a fibrillar structure resembling amyloid. It may be that the polypeptide hormones are the source of the amyloid forming in pancreatic islets, islet-cell tumours and medullary cancer of the thyroid (which secretes calcitonin). In general, however, the sites of early deposition of amyloid suggest that it is derived from a constituent of the plasma which escapes from small vessels and is possibly converted to amyloid by the digestive enzymes of phagocytic cells.

Minor components of amyloid include a second protein ('P') which constitutes about 5 per cent of both major types of amyloid. It is readily soluble and is removed from amyloid during processing of tissue; *in vitro* it forms pentamers with a ring-like structure which become stacked (like red cells in rouleaux) to form rods with periodic cross-marking. It was suggested recently that P protein is a constituent of the first component of complement but this now seems doubtful. Amyloid also contains mucopolysaccharides, lipoproteins and fibrin in trace amounts, but it is likely that these have either been trapped by amyloid as it is deposited or are plasma constituents which have permeated it.

Amyloid is readily induced in animals by prolonged antigenic stimulation or oral administration of casein, and its deposition is enhanced by thymectomy and by immunosuppressive agents. Although human amyloid is associated with conditions which involve the immunity system, there is no common immunological disturbance: for example, in lepromatous leprosy cell-mediated

immunity is depressed, whereas in tuberculosis it is normal or enhanced. This does not accord with the suggestion that amyloidosis results from T-cell depression and B-cell stimulation (Scheinberg and Cathcart, 1976).

There is some evidence that amyloid may be resorbed following effective treatment of the causal disease, but renal amyloid is persistent and steroid therapy has not, in general, proved beneficial.

Hyaline and Fibrinoid Changes

The term *hyalin** is applied to material of homogeneous, refractile, usually eosinophilic appearance seen on microscopy of stained tissue sections. It is purely descriptive, and many different formed tissue elements, as well as cell cytoplasm, may assume a hyaline appearance. In most instance the chemical basis of hyaline change is not known, although fibrin and amyloid both have a hyaline appearance. The collagen and background material of old dense fibrous tissue and the walls of aged blood vessels are often hyaline and because of its association with age this is often called *hyaline degeneration*. In the kidney and other organs, the walls of arterioles usually become thickened and hyaline in arterial hypertension (Fig. 22.6, p. 810), and glomeruli injured by chronic ischaemia became converted to hyaline balls (Fig. 22.7, p. 810). These changes occur also in diabetic nephropathy, and are believed to be due to accumulation of substances leaking out from the blood—*plasmatic vasculosis* (see below). There are many conditions in which abnormal amounts of plasma proteins leak into the glomerular filtrate; part of the protein is resorbed by the tubular epithelium where it is seen as eosinophil refractile droplets (Fig. 22.25, p. 824) commonly termed *hyaline droplets*. Protein may also coagulate in the tubular lumen and is then secreted in the urine as cylindrical '*hyaline casts*'. In virus hepatitis, damaged liver cells may appear hyaline (Fig. 20.14, p. 674) and '*Mallory's hyalin*' appears in the hepatocyte cytoplasm in alcoholic and certain other forms of liver cell injury. In Cushing's syndrome, hyaline material is seen in the basophil cells of the pituitary (*Crooke's hyaline change*—Fig. 26.4, p. 1011). Necrotic tissue, e.g. myocardium, and fused platelets in thrombi (Fig. 9.14, p. 237) may also appear hyaline.

These examples serve to show that hyaline

material and its pathological associations are widely heterogeneous. It is nevertheless sometimes of diagnostic value, as will be seen from the many examples which crop up in the systematic chapters.

A second term, **fibrinoid change**, has long been used to describe impregnation of tissues with hyaline material which is brightly eosinophilic and has other staining properties similar to those of fibrin. The term became popular in the 1940s when fibrinoid change (incorrectly regarded as diagnostic of hypersensitivity reactions) was noted to be a common feature of the so-called collagen or connective-tissue diseases. More recently, immunofluorescence staining, and to a lesser extent electron microscopy, have provided more specific techniques for identifying fibrin in tissue sections, and have shown fibrin to be present in some, but not all, 'fibrinoid' lesions: in some instances the eosinophilic hyalin is due to ground-substance mucopolysaccharides; in others its nature remains unknown.

Deposition of fibrin in the tissues results from vascular exudation of fibrinogen, which is converted to fibrin by the action of tissue thromboplastin. If the injury causing the exudation is severe, there may also be death of tissue cells, and the changes are then traditionally known as *fibrinoid necrosis*. Examples of this are seen in many acute inflammatory lesions including the Arthus reaction (Fig. 14.28, p. 381), in some infarcts (in which plasma exudes from the ischaemic blood vessels) (Figs. 2.3, p. 10 and 2.5, p. 11), in the arteriolar lesions of malignant hypertension (Fig. 14.18, p. 374) and in the necrotic base of peptic ulcers. Fibrin is detectable in some fibrinoid lesions of the connective-tissue diseases, e.g. in some examples of the subcutaneous nodules of rheumatoid arthritis (Fig. 23.53a, p. 918).

*From the Greek '*hyalos*', meaning glass.

Deposited fibrin is usually removed by the action of plasmin or by phagocytic cells. It has, however, been demonstrated in slowly developing, permanent hyaline changes in the walls of arterioles and glomeruli in plasmatic vasculosis and this supports the view of Lendrum (1969) and others that such hyaline change results from an exudative process (see also p. 809).

Corpora amylacea. Under this term are included a number of rounded or oval hyaline structures, which may stain deeply with iodine, hence the name. They sometimes show concentric lamination and may undergo calcification. Such structures form in various situations and they cannot be regarded as all of

the same nature. They are often a prominent feature within the acini of the prostate in the elderly; they occur also in the lungs, in old blood clots, and sometimes in tumours.

In the nervous system they are very common; e.g. in old age, in chronic degenerative lesions, and in the region of old infarcts and haemorrhages. They vary greatly in size, the smallest being spherical and homogeneous, and these usually stain deeply with haematoxylin. They appear to form simply by a deposition, in the intercellular spaces, of organic material containing acid mucopolysaccharides in globular form, but their exact composition is not known. They are of no importance except as a manifestation of the degenerative condition with which they are associated.

Mucins and Myxomatous Change

Mucins consist of complexes of proteins with carbohydrates and mucopolysaccharides. They are characterised by their slimy nature and histologically by their affinity for basic dyes and metachromasia with thiazine dyes such as toluidine blue. Most mucins are precipitated by acetic acid.

Mucins are secreted by various glandular epithelia and also by fibroblasts, osteoblasts and chondroblasts as important constituents of the ground substance of the various connective tissues. Both epithelial and connective tissue mucins are mixtures of *glycoproteins*, which are rich in hexose polymers and are stained pink in the PAS method, and *mucoproteins* in which the mucopolysaccharide is rich in hexosamines and which stain metachromatically with toluidine blue at low pH.

Disturbances of epithelial mucin secretion are not of much pathological importance except in *fibrocystic disease of the pancreas* (p. 721) in which an abnormality of mucus secretion occurs in the glands of the intestine, pancreas, bile ducts, bronchi and sweat glands. The thick mucin secreted obstructs the ducts, with subsequent gland atrophy and loss of function.

Obstruction of the ducts of small mucus-secreting glands, e.g. in the mouth, results in the development of mucin-filled cysts, and obstruction of the cystic duct may result in distension of the gallbladder with mucin—the so-called *mucocoele* (Fig. 20.67, p. 711). Chronic irritation of a mucous membrane may result in

increase in the number and activity of mucin-secreting cells, as in chronic bronchitis in which there is abundant mucous sputum. Some epithelial tumours secrete mucin, and its detection in relation to tumour cells is sometimes of help in determining the origin of the tumour.

The mucopolysaccharides of **connective tissue mucins** include hyaluronic acid, chondroitin, chondroitin sulphates and other sulphated compounds. They form the ground substances of fibrous tissue, cartilage and bone, and also joint fluid. In the soft tissues, the ground substance is largely in the form of a gel, but in acute inflammatory lesions the mucopolysaccharides are depolymerised, with conversion mainly to a fluid phase, which is more readily permeable to exudate and cells of the inflammatory reaction. Some bacteria also secrete hyaluronidase and other enzymes which may facilitate their spread in the tissues.

Some connective tissue tumours secrete abundant mucin, which appears as a basophilic stroma; they are called *myxomas* (Fig. 13.6, p. 343) and an increase in mucoid ground substance of connective tissue, so that it comes to resemble myxoid tissue of the fetus and umbilical cord, is termed **myxomatous** or **myxoid change**. It occurs in the aortic media in *Erdheim's medial degeneration* (Fig. 14.34, p. 386), and also, together with similar changes in other connective tissues, in *Marfan's syndrome*: in both conditions the inner part of the weakened aortic wall may rupture and blood may

track along the media (dissecting aneurysm). Myxomatous change is also seen in the valve cusps of the heart and sometimes results in stretching and incompetence of the valves.

Production of ground substance is influenced by hormones; there is a generalised increase, for example, in *hypothyroidism* giving rise to the term **myxoedema**: the bloated appearance of the face is due to myxomatous change in the dermis, and the croaky voice is due to the same change in the larynx. Curiously, myxomatous change is seen in the pre-tibial region in some cases of thyrotoxicosis (hyperthyroidism).

There are also a number of defects of mucopolysaccharide metabolism—the **mucopolysaccharidoses**—which are inherited as Mendelian recessive characters. Excess mucopolysaccharides accumulate in various types of cell, and are excreted in the urine. The conditions are distinguished by the chemical nature and distribution of the material. The best known example is *Hurler's syndrome* or *gargoylism*, the major features of which include dwarfism, skeletal deformities, a characteristic facies, mental deficiency, corneal opacities and hepatomegaly.

Melanin Pigmentation

The melanins are iron-free sulphur-containing pigments varying in colour from pale yellow to deep brown. They are formed intracellularly from colourless precursors—melanogens—and are very stable substances, resistant to acids and many other reagents, but soluble in strong alkalis; they can be bleached by powerful oxidising agents such as potassium permanganate or hydrogen peroxide. They are related to the aromatic compounds, tyrosine, phenylalanine and tryptophane and may be formed from such substances by oxidation. On treating sections of skin with dihydroxyphenylalanine (dopa), 'dopa-positive' cells in the epidermis oxidise this substance by means of an enzyme like tyrosinase and become blackened in consequence. The only cells in the skin which are 'dopa-positive' *in vivo* are the *dendritic cells*, or *melanocytes*, which lie extended between the basal cells of the epidermis (Fig. 10.6); they are the only melanin-producing cells in the skin and fine granules of melanin in their dendrites are taken up by pinocytosis of the tips of the dendrites, into adjacent epidermal cells and also into certain phagocytic cells (*melanophores*) in the dermis which may thus become heavily laden with coarse pigment granules. Melanin granules possess the capacity to reduce certain silver salts, e.g. ammoniacal silver nitrate, with consequent deposition of metallic silver; melanin can thus be blackened in histological preparations, scanty or light-coloured granules being rendered conspicuous. This property is widely used histochemically. It is now known that the dendritic cells are of neuro-ectodermal origin, being derived from the cells of the embryonic neural crest, as are also the melanocytes of the squamous mucous membranes, the meninges, choroid and adrenals. This view harmonises well with the evidence about the origin of the naevus cells of pigmented moles from neuro-ectodermal cells and also accords with the



Fig. 10.6 Dendritic cells (melanocytes) in the basal part of the epidermis. (Dopa reaction.) $\times 220$.

experimental work of Billingham and Medawar on the behaviour of melanocytes in skin autotransplants. This work seems to have rendered untenable the alternative view that melanocytes are modified basal epidermal cells. The Langerhans cells of the epidermis are now regarded as macrophages and may play a part in the control of keratinisation. Darkening of the skin on exposure to ultraviolet radiation is brought about first by migration of the melanin granules and subsequent darkening of their colour; later there is increased formation of pigment,

apparently by the activity of the dendritic cells, which under further stimulation may increase in number.

In **Addison's disease**, which results from destruction of the adrenal cortex (p. 1044), there occurs a general increase in melanin pigmentation of the skin, especially in areas exposed to light and in areas normally pigmented. There may also be pigmentary deposition on the inner surface of the cheeks on a line corresponding to the junction of the teeth, and on the sides of the tongue, the position being apparently determined by irritation. In the skin the pigment is in the form of very fine brownish granules in the deeper layers of the rete Malpighii, and is present also as coarser granules, chiefly within macrophages in the underlying cutis, the appearance and distribution resembling those in the negro skin. The pigmentation in Addison's disease represents an increase of normal pigment, and occurs under the influence of the melanocyte-stimulating hormone of the pituitary (MSH) which is released in excess in the absence of adrenal inhibition (p. 1044).

Chloasma is a condition observed principally during pregnancy, and occasionally in association with ovarian disease, in which pigmented patches occur in the skin of the face, and the pigmented parts, e.g. the nipples, may become darker under the influence of oestrogenic and melanocyte-stimulating hormones. A similar condition has been described in women taking oral contraceptives.

Leukoderma (vitiligo) denotes patchy depigmentation of skin and this may be accompanied by increase of pigment in the intervening areas. In the affected areas the dendritic cells are of abnormal structure and have lost their capacity to oxidise dopa to form pigment.

Irregular pigmentation of the skin is common in chronic arsenical poisoning and in neurofibromatosis. In haemochromatosis also, the colour of the skin is due partly to deposition of haemosiderin in the cutis, notably around the sweat glands (p. 282), but also to increase in melanin. A striking degree of melanotic pigmentation of the oral and labial mucosa occurs in association with familial multiple polyposis of the small intestine, especially the jejunum (Peutz-Jeghers syndrome): the disorder is transmitted as a Mendelian dominant. The control of pigment metabolism in the skin is obscure, but it is known to be affected by exposure to

light, chronic irritation and increased vascularity, activity of endocrine glands including the adrenals, pituitary and ovaries, and nervous influences.

Pigmented tumours. Melanin pigment is formed in large amount in the melanotic tumours which arise in the skin and in the pigmented coats of the eye, and most analyses have been carried out on the pigment from such tumours. The urine of patients suffering from extensive melanotic tumours occasionally contains a melanogen which darkens on exposure to the oxygen of the air.

Melanosis coli. This is a rather uncommon condition characterised by varying degrees of brownish to black pigmentation of the mucosa of the colon, beginning in the caecum and ascending colon, and sometimes extending to the anus. The pigment is contained mainly in macrophages in the lamina propria; it is absent from the epithelial cells. The condition is commonest when there has been intestinal stasis or chronic obstruction, and it is now recognised to be the result of absorption of aromatic products from the gut. This is commonly associated with the prolonged use of anthracene-derived purgatives, e.g. cascara, and the pigment consists of derivatives of anthraquinone combined with products of protein decomposition. The pigment resembles melanin in its reactions, but differs from it in being autofluorescent, weakly PAS-positive, and weakly sudanophilic. The cells containing pigment are dopa-negative.

Ochronosis. In this very rare condition, cartilages, capsules of joints and other soft tissues assume a dark brown or almost black colour, owing to pigment deposition. The pigment resembles melanin in some of its properties but does not reduce silver nitrate. In virtually all cases of ochronosis, alkaptonuria is present, a condition in which homogentisic acid (2,5-hydroxyphenylacetic acid) is excreted by the kidneys and causes the urine to blacken on standing owing to oxidation, especially alkaline urine. Homogentisic acid is formed from tyrosine and phenylalanine. Normally it is converted to malyl-acetoacetic acid by homogentisic acid oxidase in the liver and kidneys, but alkaptonurics lack this enzyme, and consequently homogentisic acid is not metabolised normally, but is oxidised into pigment and deposited in the tissues, producing ochronosis. The metabolic defect in alkaptonuria is inherited as an autosomal recessive character. In the early days of antiseptic surgery ochronosis occasionally followed the use of carbolic dressings for a long time, and the pigment is believed to be formed from the absorbed carbolic acid. This has been called *exogenous ochronosis*.

Pigments Derived from Haemoglobin

At the end of their life span, red cells are taken up by macrophages in the spleen, marrow, etc. Intracellular breakdown of haemoglobin (Hb) begins with opening of the porphyrin system of haem, the four pyrrole nuclei and globin now forming a long-chain molecule (choleoglobin). The globin and iron are then split off and the residual *biliverdin* pigment, consisting of four pyrrole rings, is reduced to *bilirubin* and passes into the plasma where it is bound mainly to albumin. The bilirubin is taken up by the hepatocytes, dissociated from the protein, and is conjugated with glucuronic acid and excreted as bilirubin glucuronides in the bile. The iron which is split off from haem is stored mainly as *ferritin* and *haemosiderin* and re-used.

Breakdown products of Hb may accumulate in the body in the following circumstances: (a) local deposition results from haemorrhage into the tissues; (b) more generalised accumulation of bilirubin occurs when there is excessive red cell destruction, i.e. in haemolytic anaemias; (c) increase in bilirubin or its glucuronides occurs when there is some defect in the metabolic or excretory pathways by which the iron-free part of haem is delivered into the intestine as bilirubin glucuronide; (d) accumulation of iron-containing compounds occurs when the amount of iron entering the body exceeds significantly the small amount which is lost physiologically. The effects of these abnormalities are described below.

Local accumulation of pigments

When haemorrhage into tissues occurs, many of the red cells in the escaped blood undergo lysis; their Hb diffuses away and is taken up and catabolised in macrophages in the draining lymph nodes, spleen, etc. However, some of the red cells are phagocytosed locally by macrophages derived from monocytes which migrate into the lesion and bilirubin and iron compounds are produced as described above: this process is illustrated experimentally in Figs. 10.7 and 10.8: in man it is reflected in the changing colours of a 'black eye' or any other superficial bruise. Most of the bilirubin diffuses away and is eventually dealt with by the liver, but some of it may persist locally in crystalline

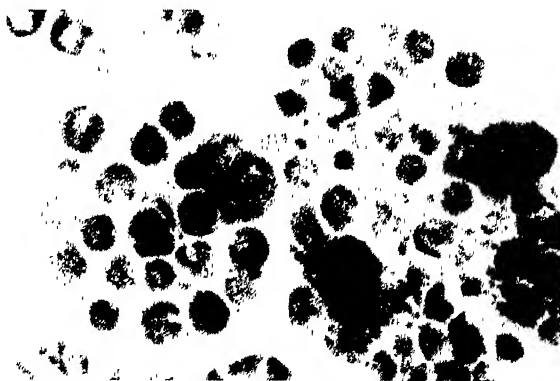


Fig. 10.7 Macrophages containing red cells in phagocytic vacuoles. From the subcutaneous tissue of a mouse six days after an injection of red cells. $\times 1200$.

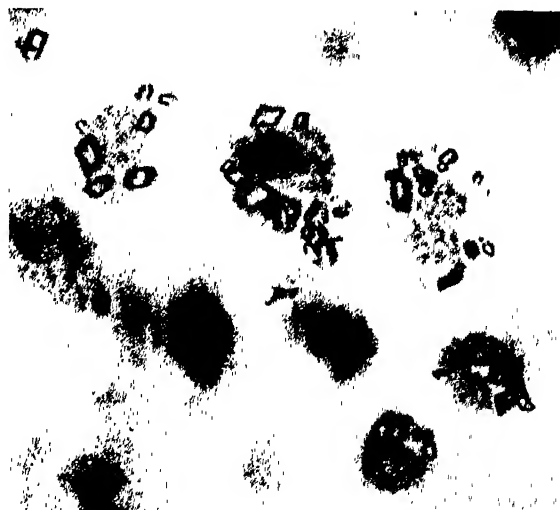


Fig. 10.8 Intracellular formation of bilirubin crystals in macrophages of mouse, 16 days after injection of haemoglobin. $\times 1200$. (From preparations by the late Dr. Janet S. F. Niven.)

form around an old haemorrhage, particularly in the brain (Fig. 10.9): this may be due to the absence of lymphatics in brain tissue. Some of the iron released may also be retained locally as the pigment *haemosiderin* (see below), either within macrophages or as an incrustation of collagen and other tissue components.

Localised accumulation of haemosiderin may occur in the **lungs** as a result of haemorrhages in pulmonary venous congestion, e.g. in *mitral stenosis* (Fig. 10.10) and also in *idiopathic pulmonary haemosiderosis* where it is accompanied by fibrosis. Haemosiderin deposition also



Fig. 10.9 Crystals of bilirubin and granular pigment, some of which is in phagocytes, at the site of an old cerebral haemorrhage. $\times 500$.

occurs in the **renal tubular epithelium** when intravascular haemolysis results in release into the plasma of haemoglobin, which leaks into the glomerular filtrate and is taken up by the tubular cells (p. 525).

Bile pigments

A rise in the level of **bilirubin** in the plasma results from increased breakdown of red cells in the haemolytic anaemias, or from failure of the liver cells to remove and conjugate it with glucuronic acid. Lesions of the liver or biliary tract which prevent excretion of **bilirubin glucuronide** result in its regurgitation into the plasma. When the levels of either compound exceed 2–3 mg per 100 ml (35–50 $\mu\text{mol/l}$), **jaundice** develops, i.e. the skin, sclera and various other



Fig. 10.10 The lung in mitral stenosis. Red cells escaping from the congested pulmonary capillaries are ingested by alveolar macrophages which became engorged with haemosiderin. The macrophages accumulate in the alveoli adjacent to respiratory bronchioles and are thus seen as aggregates. Prussian blue reaction. $\times 50$.

tissues become distinctly yellow. In adults, jaundice itself causes little disability, but in infants a rise of (unconjugated) bilirubin in the plasma to over 15 mg per 100 ml (250 $\mu\text{mol/l}$) carries a risk of toxic brain injury. The types of jaundice and their causes and effects are, however, dealt with more fully in Chapter 19.

Iron pigments

About 70 per cent of the 3–4 g of iron in the body is incorporated in the haem of haemoglobin: 5 per cent is in myoglobin and small amounts are incorporated in cellular cytochrome, respiratory and metallo-flavo enzymes. The remainder (1–1.5 g) is mostly in storage form in macrophages of the spleen, bone marrow, etc. and in various tissue cells, but particularly hepatocytes.

Iron absorption is by way of the epithelial cells lining the gut, mainly those of the villi of the duodenum and proximal jejunum, absorption decreasing progressively more distally. Dietary iron consists of both haem (from meat and fish) which is absorbed as metalloporphyrin, and non-haem iron which is absorbed mainly as ferrous salts, ferric salts being poorly absorbed. The cells of the villi probably

possess receptors for haem and for ferrous salts, both of which appear to be taken into the cells by endocytosis. The amounts of iron taken up by the epithelium depend mainly on two factors, the body's total storage iron and the amount of available iron in the diet. *The amount taken up is inversely proportional to the total storage iron in the body* (Fig. 10.11) and

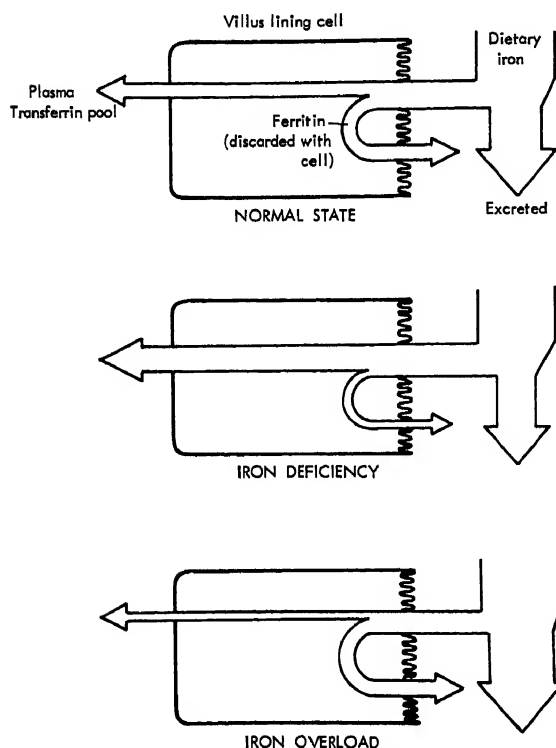


Fig. 10.11 The effect of the total iron store on absorption of dietary iron by intestinal epithelium. Only a fraction of the available dietary iron is normally taken up by the epithelium (*upper*): the fraction taken up is increased when total storage iron is depleted (*middle*) and decreased when the iron store is increased (*lower*). Of the iron taken up by the epithelium, the proportion transferred to the plasma transferrin pool is similarly affected by the total iron store.

directly proportional, within limits, to the amount available in the diet. It is not known how the storage iron influences absorption, but it seems likely that in iron deficiency states the number of iron receptors on the lining cells of the villi is increased. These observations apply to absorption of both haem iron and ferrous salts. Other dietary factors also influence iron absorption. For example, absorption of **non-haem iron** is aided by ascorbic acid, citric acid

and amino acids, all of which form monomeric complexes with iron and prevent the formation of non-absorbable polymers. Gastric HCl also favours absorption by preventing polymer formation and alcoholic drinks increase iron intake, possibly by stimulating gastric secretion but also because some drinks are rich in iron. Reduction of ferric to ferrous iron by ascorbic acid and other reducing agents also promotes absorption of non-haem iron. Formation of non-absorbable polymers is favoured by gastric achlorhydria and by the presence in the diet of certain compounds, e.g. phytate from cereals, tannates, calcium phosphate and ethylenediamine tetra-acetic acid (EDTA) used as a food preservative. These factors explain why iron deficiency is common among people living on a largely vegetable diet. **Haem** is a particularly important source of iron because, although it represents only a small part of the total dietary iron, the proportion absorbed is relatively high: its absorption is favoured by amino acids and is not inhibited by dietary factors (or achlorhydria) which inhibit absorption of non-haem iron.

Within the mucosal epithelial cells, haem iron is broken down and forms a common pool with the absorbed non-haem iron. Part of this pool is bound to a transferrin-like protein and is rapidly transferred through the epithelial cells to enter the plasma. The remainder is incorporated into intracellular ferritin (see below), much of which is lost when, within a few days, the cell exfoliates. The amount of epithelial-cell iron transferred to the plasma increases with the concentration of iron in the lumen of the gut, but the *proportion* transferred diminishes. Over a wide range of iron concentrations in the lumen, *the proportion of iron transferred is greater in subjects with iron deficiency, i.e. with depleted iron stores*, and is diminished when the iron stores are large (Fig. 10.11). How these factors influence the proportions of epithelial iron which bind to the transfer protein and to ferritin is not known.

Iron loss amounts to approximately 1 mg daily, most of it in the form of ferritin in the epithelial cells desquamated from the skin (0.2–0.3 mg) and gut (0.6 mg). Only a small proportion is lost in the bile, sweat, etc., and some of that lost by gut epithelium has been absorbed from the lumen by the cells, and incorporated into ferritin; it has never really entered the

body's iron pool. In women of reproductive age, menstruation accounts for an average loss of an additional 0.6 mg daily, although there is enormous individual variation. Pregnancy, childbirth and lactation represent a rather greater loss than this.

From the above it will be apparent that *iron balance depends largely on the control of absorption from the gut: iron loss is small and subject to much less variation than is absorption.*

Plasma iron. The plasma of normal adults contains an average of about 120 μg of iron per dl, although the range is large. Over 95 per cent of this is in the form of **transferrin**, which consists of ferric iron bound to a specific transport protein, a β -globulin termed **apotransferrin**. This possesses two binding sites for iron and is normally only about 30 per cent saturated in the plasma. *Although transferrin makes up only a very small percentage of total body iron, it is very important, for it is the form in which iron is transferred from macrophages, parenchymal cells and intestinal epithelium to the erythropoietic cells.* In fact, the plasma iron is provided mainly by macrophages of the spleen, haemopoietic marrow, liver, etc., which break down red cells and synthesise apoferritin: these cells thus release transferrin, most of which is taken up by red cell precursors and used for haemoglobin synthesis.

A small proportion of plasma iron is in the form of ferritin (see below) and this is of importance because its concentration reflects the size of the total iron store of the body.

Haemoglobin synthesis. Red cell precursors have surface receptors for transferrin, which they take up avidly, probably by endocytosis. Uptake is proportional to the concentration in the plasma. Within the cell, iron is split off and used in haem synthesis, while the apotransferrin is returned to the plasma. With increasing maturity, the number of transferrin receptors on the erythroid cell diminishes, the mature red cell having none.

Storage iron. Iron is stored within cells in two forms, ferritin and haemosiderin. Most if not all cells synthesise apoferritins and store iron as **ferritin**, which consists of micelles of ferric oxide phosphate enclosed in a protein molecule which is water soluble. Ferritin is capable of incorporating approximately 5000 atoms of iron per molecule, but it is never fully saturated and provides an immediate reserve

iron storage capacity: it is not detectable by light microscopy but has a characteristic electron-microscopic appearance. The amount of iron stored as ferritin is limited and normally most of the total iron is stored in more concentrated form as **haemosiderin**. This is probably formed from ferritin and has a ferric iron content of up to 40 per cent. It is insoluble and if present in large amount is seen microscopically as golden-yellow intracytoplasmic granules, while it imparts to the tissue a brown appearance to the naked eye. Haemosiderin gives the prussian blue reaction on treatment with hydrochloric acid and potassium ferrocyanide, the intense blue colour being due to formation of ferri-ferrocyanide.

Most of the stored iron is present in the form of haemosiderin in macrophages in the spleen, marrow, etc. In parenchymal cells, notably in the liver, iron is normally stored mainly as ferritin, although there may be sufficient haemosiderin in the hepatocytes to give a faint prussian blue reaction. Absence of microscopically detectable haemosiderin in the macrophages in smears or sections of haemopoietic marrow indicates depletion of the iron stores. In states of increased storage, the proportion of iron in the form of haemosiderin increases and in gross iron overload it may be present in enormous amounts, its distribution between macrophages and parenchymal cells depending on the cause of the iron overload (see below).

In states of negative iron balance, iron is transferred from intracellular ferritin and haemosiderin to the plasma transferrin pool, most of it being provided by macrophages and hepatocytes, and anaemia does not develop until after the stores are depleted.

Iron deficiency

This is the commonest disturbance of iron metabolism. It results first in depletion of storage iron, secondly in anaemia, and lastly, and less certainly, in fall of the cytochrome content of cells. Most of the important effects of iron deficiency are due to anaemia, and accordingly the subject is dealt with in relation to the blood in Chapter 17.

Iron overload

This can result either from absorption of excessive amounts of iron from the gut or from

administration of parenteral iron, for example by multiple blood transfusions.

The outstanding example of naturally occurring iron overload is provided by the disease termed **idiopathic haemochromatosis**, in which storage occurs predominantly in the parenchymal cells of the internal organs where it has serious consequences. By contrast, when multiple transfusions are administered over a period of years to patients with aplastic anaemia (due to marrow aplasia), iron accumulates mainly in macrophages, where it causes little injury. The effects of iron overload thus depend not so much on the amount of iron stored, as on its distribution between macrophages and parenchymal cells.

In iron overload due to excessive dietary iron or oral iron therapy, and in certain disorders of haemoglobin or red cell production, the distribution of stored iron, as explained later, is more complex.

States of increased iron storage are conveniently termed **haemosiderosis** or **siderosis**.

Idiopathic haemochromatosis

This is characterised by excessive absorption of dietary iron, probably from birth. The total iron of the body gradually increases until, by the age of 40 years or so, it may exceed 20 g instead of the normal 3–4 g. Erythropoiesis is normal and the excess of iron is stored as haemosiderin mainly in the parenchymal cells, particularly in the liver (Fig. 10.12), pancreas and myocardium, but also in many other organs.

The condition usually affects men, and the quantity of iron stored depends on the amount of available iron in the diet and is also increased by heavy alcohol consumption, which is common among affected individuals. Iron in excess has a cytotoxic effect: hepatocyte destruction with accompanying fibrosis leads to cirrhosis, which is often the cause of presenting symptoms. The toxic effect on the myocardial cells commonly leads to congestive heart failure and cardiac arrhythmias, while heavy deposition in the pancreas results in cell loss and fibrosis. Over 50 per cent of patients develop diabetes, probably due to a combination of liver and islet-cell injury. The liver, pancreas, etc. appear brown to the naked eye and give an intense prussian blue reaction. Microscopy shows heavy deposition of haemosiderin in the

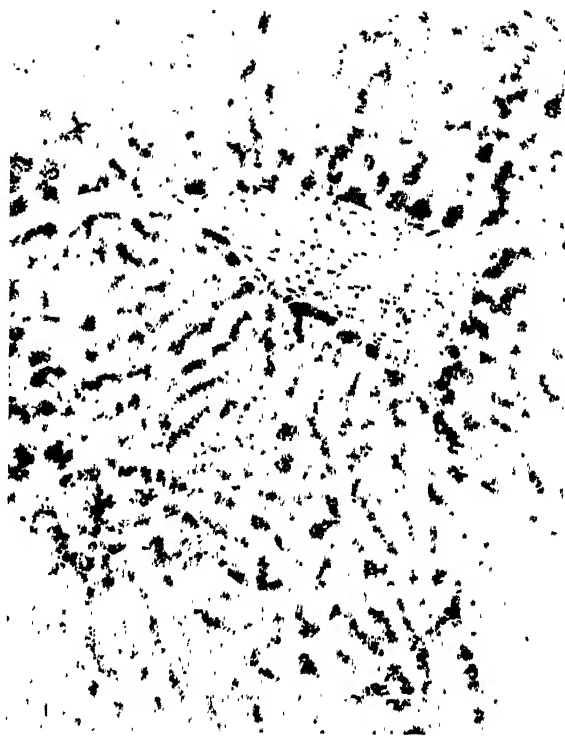


Fig. 10.12 Needle biopsy of the liver in haemochromatosis. The excess iron is stored mainly as haemosiderin in the hepatocytes, and is seen as dark granules. Prussian blue reaction. $\times 150$.

parenchymal cells. Death of damaged cells results in release of haemosiderin, which is therefore seen also in the adjacent stroma and in macrophages in the affected organs and their draining lymph nodes. Apart from this, the amount of haemosiderin in macrophages in general, e.g. in the spleen and bone marrow, is not greatly increased. The gastric mucosal cells are rich in haemosiderin, but not the cells of the villi of the upper small intestine. The skin develops a bronzed appearance (hence the term **bronzed diabetes**) due to excess melanin production, which is unexplained; in some cases, however, the skin appears more leaden owing to iron deposition, mainly in relation to the sweat glands. Other features include the polyarthritides of pseudo-gout, due to the formation in the joint tissues and spaces of calcium pyrophosphate crystals; this may result from the inhibitory effects of iron on pyrophosphatase. Hypogonadism, probably due to injury to the adeno-hypophysis, is common, and also vague neurological symptoms of unknown cause.

Aetiology. The nature of the metabolic defect responsible for excessive iron absorption is

unknown. The plasma transferrin is not increased but is fully saturated with iron (cf. normal 30 per cent saturation). Tests of iron absorption have given conflicting results, but if the iron stores are depleted by venesection etc., absorption can then be demonstrated to be increased. If the iron stores are allowed to re-accumulate, absorption falls, sometimes to normal. The inverse relationship between total iron store and absorption is thus maintained, but *the proportion of dietary iron absorbed at all levels of iron storage is abnormally high*: as noted by Bothwell *et al.* (1979) the 'absorbo-stat' is set too high.

Although the nature of the defect is unknown, it appears that macrophages are incapable of storing excess iron as haemosiderin, and that in consequence the plasma transferrin becomes saturated and so excess iron is taken up by the parenchymal cells of the liver, etc. Because transferrin is saturated, iron absorbed from the gut and transferred to the plasma may be transported by the portal circulation in a form which is readily taken up by the liver cells, thus explaining why the liver is particularly severely affected.

Inheritance of haemochromatosis has been much debated. Occasionally it affects more than one member of a family, and investigation of the relatives of patients has shown that many of them, although apparently healthy, have a sub-clinical form of the disease with lesser degrees of increased iron storage, increased saturation of plasma transferrin, and sometimes liver injury short of cirrhosis. Such individuals are believed to be particularly prone to develop the clinical picture of haemochromatosis if their alcohol consumption is high. It has been suggested that the disease is inherited as a recessive character, the homozygous state resulting in overt haemochromatosis and the heterozygous state causing sub-clinical iron overload. In both the overt and sub-clinical disease, alcohol probably acts by its toxic effect on the liver and by increasing iron absorption.

The importance of genetic factors is suggested also by the reported high incidence of HLA-A3 and B7 antigens in patients, and by the occurrence of the sub-clinical form of the disease in siblings of identical HLA types. The nature of genetic inheritance is, however, unlikely to be elucidated until a test for the defect in the presumed heterozygote becomes available.

Apart from treatment of heart failure, diabetes etc., reduction of the excess iron store, for example by repeated phlebotomy and by limit-

ing the dietary iron, is the most effective form of therapy.

Iron overload in anaemia

Iron deficiency is a common cause of anaemia, but in certain types of anaemia due to other causes there may be greatly increased iron storage. For example, in aplastic anaemia, in which haemopoiesis fails and the marrow becomes hypocellular, life can be maintained only by regular blood transfusions and since each unit of blood contains 200–250 mg of iron, gross iron overload can develop over a number of years. In contrast to haemochromatosis, most of the iron accumulating as haemosiderin is stored in macrophages in the spleen, liver (Fig. 10.13), marrow and elsewhere. There is some increase of iron in the hepatocytes; initially it is relatively slight, but later may increase as a result of redistribution of iron.

The situation is different in patients in whom anaemia is due to defective red cell production

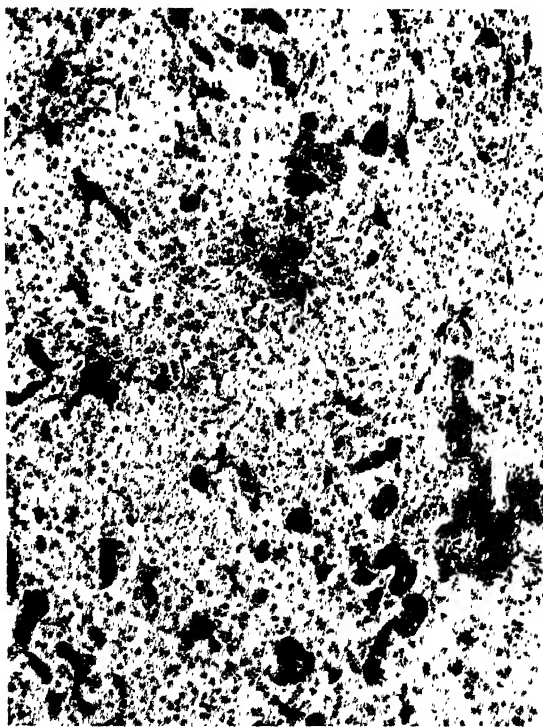


Fig. 10.13 Needle biopsy of the liver in a case of haemosiderosis resulting from multiple blood transfusions. Haemosiderin is present in groups of enlarged Kupffer cells and macrophages in the portal areas. The patient suffered from chronic renal failure and had been maintained on haemodialysis. $\times 150$.

in spite of hyperplasia of the erythropoietic tissue, i.e. in which there is '*ineffective erythropoiesis*'. A good example is provided by thalassaemia major (p. 524) in which there is defective haemoglobin synthesis, and in the familial form of sideroblastic anaemia (p. 540) in which incorporation of iron into haemoglobin is defective. In such conditions, immature erythroid cells are destroyed in the marrow. In some unknown manner, increased but ineffective erythropoiesis increases iron absorption and thus causes overload, which, as in haemochromatosis, results in deposition in parenchymal cells. If the condition is inherited and so present from birth onwards, a picture similar to haemochromatosis may develop.

In anaemia with increased but *effective* erythropoiesis, e.g. chronic haemolytic anaemia, in which circulating red cells are destroyed abnormally rapidly, and in which compensatory increase in erythropoiesis occurs, there is little or no increase in iron absorption and overloading does not usually occur unless multiple transfusions are administered.

Dietary iron overload

It is rare for serious iron overload to occur from increased iron intake in the diet, partly because in most diets with a high iron content much of the iron is in unavailable form, and partly because any increase in iron stores inhibits absorption of dietary iron. There is, however, one outstanding example of overload resulting from dietary factors, and that is in South Africans of the Bantu tribe. Their intake of iron is very high, partly because iron pots are used in cooking, but mainly from drinking beer brewed in iron containers. Because of its low pH, the brew dissolves iron and as much as 50–100 mg of iron may be ingested daily in a few litres of the (rather weak) beer. The distribution of iron varies: in many cases, however, it accumulates both in macrophages and in the parenchymal cells in the liver and sometimes other organs, and the picture of haemochromatosis with hepatic cirrhosis and sometimes diabetes may develop. In such cases, the degree of

liver injury has been shown to correlate partly with the amount of iron in the hepatocytes. A striking difference from haemochromatosis is the deposition of haemosiderin in the lamina propria of the villi of the proximal small intestine. In other individuals, storage of iron is predominantly in macrophages and without serious effects. It is not known why iron deposition is parenchymal in some and in macrophages in others, but it has been reported that, in the former, there is a high degree of saturation of plasma transferrin. With the increasing use of commercially prepared beverages, the incidence of haemosiderosis in the Bantu has declined considerably. The alcohol in the beer also doubtless contributes to the hepatic injury and in some cases the picture is that of alcoholic cirrhosis with excess of iron deposition.

The risk of serious iron overload from prolonged taking of medicinal iron by mouth appears to be slight, for although there are reports of haemochromatosis developing, in many other cases the iron stores do not appear to have been very greatly increased. Acute iron poisoning can, however, result from gross over-dosage with iron.

Malarial pigmentation

In malaria the parasites within the red cells produce from the haemoglobin a dark brown pigment, haematin, in the form of very minute granules, which accumulates within the parasites. When the adult divides into young forms (merozoites), the red cell disintegrates and the pigment is released to be taken up by monocytes (Figs. 17.22, 17.24, p. 529) and by macrophages, especially in the spleen, liver and haemopoietic marrow, where it remains practically unchanged for many years. In chronic malaria these tissues appear dark brown. Malaria pigment does not give the prussian blue reaction and resembles closely the artefact pigment derived from formalin acting on blood. When there is much blood destruction, especially in severe cases of malaria, haemosiderin may be deposited in the organs in addition to the malarial pigment.

Lipofuscin: Age Pigment

In the later years of life a fine brownish-yellow pigment tends to appear in the heart muscle, smooth muscle, etc.; and in wasting diseases this accumulation of pigment is more marked. In some cases of malabsorption syndrome, for example due to coeliac disease, it is present in the smooth muscle of the small intestine and oesophagus, and in smaller amounts in that of the stomach and colon: experimental studies suggest that vitamin E deficiency may be responsible.

In the heart muscle the pigment accumulates in the central part of the cells around the poles of the nucleus, and when this is associated with wasting of the muscle, the term *brown atrophy* is applied. Similar pigment may occur in the liver cells, especially in the central parts of the lobules, in the cells of the testis, and in the nerve cells of the cortex of the brain. Heavy deposits of pigment in the cortical neurons are

seen in senile dementia and allied conditions. The pigment must be distinguished from that which occurs normally in the pigmented neurons of the locus caeruleus and substantia nigra, which belongs to the melanin group. In brown atrophy the pigment is believed to be chiefly lipid, as it reduces perosmic acid and is usually coloured by the sudan stains. It is often called *lipofuscin*, but differs in its chemical and staining reactions in the various organs, some being fluorescent, doubly refracting or acid-fast in varying degree, e.g. *ceroid*, an acid-fast pigment found in the liver in certain forms of experimental cirrhosis. In electron micrographs it is seen as *residual bodies* (p. 23), which result from incorporation of cell constituents into phagosomes in the process of autophagocytosis, the lipofuscin persisting as indigestible residues of cellular lipids.

Exogenous Pigmentation

Inhaled compounds. The most important exogenous pigments are those inhaled as dust particles and entering the body through the respiratory passages. A certain amount of soot, stone dust, etc., enters and accumulates in the lungs of all individuals living in urban conditions, but the accumulation becomes excessive in those exposed occupationally to an atmosphere rich in dust. The lungs may be infiltrated by foreign particles of various kinds—coal, silica, asbestos, iron and other ores and various organic substances. The resulting pathological changes will be described later with the diseases of the lungs.

The entrance of such particles into the lungs is favoured by the presence of chronic bronchitis or other condition in which there is interference with the action of the ciliated epithelium, but even in normal health, particles of less than $5\text{ }\mu\text{m}$ gain access to the pulmonary alveoli if the amount in the inspired air is great. The dust particles are quickly taken up by macrophages in the pulmonary alveoli (Fig. 10.14). Some of the macrophages with the ingested particles are expelled *via* the bronchi, some enter



Fig. 10 14 The lung of a coal-miner showing phagocytosis of inhaled particles of coal dust by alveolar macrophages. $\times 520$.

the interstitial tissue of the lungs, and pass into the lymphatics, while some settle in the alveoli alongside respiratory bronchioles, and the pigment is eventually incorporated into the respiratory bronchiolar walls. Much of the pigment, however, is carried into the lymphatics; most of it is deposited in the hilar nodes, but some in the pleura. The degree of irritation resulting depends on the nature of the particles. Large collections of carbonaceous particles (*anthracosis*) may provoke little or no overgrowth of connective tissue, whereas fibrosis is very marked in the case of silica-containing stone dust, the condition of *silicosis* resulting. The bronchial lymph nodes become pigmented and enlarged, the accumulation within their phagocytic cells being virtually permanent. Some of the pigment which has accumulated in the lungs may be removed by macrophages which appear in the sputum for a long time after removal of the individual from the dusty atmosphere.

Ingested compounds. Deposition of brownish granules of silver compounds (*argyria*) was a

common result of taking medicines containing silver preparations. The granules are formed by reduction of silver albuminate and are seen especially in the skin (giving a dusky appearance), the gut wall, and the basement membranes of the glomeruli and renal collecting tubules. It is now rare. In *chronic lead-poisoning* an albuminate is produced in a similar way, and around the teeth hydrogen sulphide reacts with it to produce the characteristic blue line on the gums. *Melanosis coli* (p. 277) is now the commonest example of pigmentation resulting from ingestion of chemicals.

Tattooing. In tattooing, fine particles such as india ink, ultramarine, cinnabar (mercuric sulphide), etc., introduced through the epidermis, are taken up by macrophages and lodge in small spaces or clefts in the connective tissue of the cutis. Some particles are carried also by the lymph stream to the regional lymph nodes and then are conveyed by phagocytes into the lymphoid tissue. Both at the site of introduction and in the lymph nodes the pigment persists for life.

Pathological Calcification

Pathological calcification of soft tissues occurs most commonly without any general disturbance of calcium metabolism: the level of plasma calcium is normal, and deposition is due to local changes in the affected tissue. This is termed *dystrophic calcification*. Less commonly, pathological calcification is a result of an increase in the level of ionic calcium in the plasma, and occurs in normal soft tissues: this is termed *metastatic calcification*.

In both dystrophic and metastatic calcification the deposits resemble in composition the minerals of bone, but show much greater variations in the proportions of calcium to magnesium and phosphate to carbonate.

Identification of calcium salts in tissues. Calcium salts have an affinity for haematoxylin, and the earliest sign of calcification is given by the appearance of hyaline or finely granular material of a deep violet tint. Later the calcium salts form irregular and somewhat refractile masses: they are, of course, readily soluble in weak acids, and small bubbles

of carbon dioxide are released from the carbonates. When treated with dilute sulphuric acid, the characteristic crystals of calcium sulphate separate out. This occurs more readily when the sections are in 50 per cent alcohol, in which the solubility of the crystals is low. When carbonate or phosphate (which are nearly always deposited as calcium salts) are treated with silver nitrate, yellow silver phosphate is formed, and this quickly undergoes reduction on exposure to light and turns black (von Kossa's method). Neither the affinity for haematoxylin nor von Kossa's method is specific for calcium. Silver nitrate is reduced by other substances, e.g. iron, and the reaction with haematoxylin is given by a substance formed before the deposition of calcium, and is positive after the tissue is decalcified. The best reagent is alizarin, the staining principle in madder, or its derivatives. Alizarin stains calcium salts red, but the reaction may not be given by very old deposits. When injected *intra vitam*, alizarin colours growing bone (but not fully formed bone) and also pathological deposits of calcium unless they are very old.

Calcification is often accompanied by diffuse or granular deposition of iron compounds which give a prussian blue reaction.

Dystrophic calcification

This consists of the irregular deposition of calcium salts in altered or necrotic tissues and formed elements such as thrombi. Deposition is irregular and may be sufficiently heavy to render the part chalky or even stony hard.

Predisposing changes. The local changes which predispose to dystrophic calcification are as follows.

(1) *Hyaline changes in fibrous tissue.* This occurs as an ageing change in arteries. Increase in calcium in hyalinised artery walls is usual, and it may be sufficient to convert the vessel to a rigid tube, as in Monckeberg's sclerosis (Fig. 14.19, p. 376). Calcification is also common in dense connective tissues, for example tendons, the dura mater, and the scarred heart valves following rheumatic endocarditis. It occurs in some tumours, for example in fibromas (Fig. 10.15) and in uterine myomas undergoing involution after the menopause. The 'brain-sand' bodies of some meningiomas consist of concentrically arranged cells which undergo hyaline change followed by calcification.



Fig. 10.15 Dystrophic calcification of hyaline connective tissue adjacent to a small blood vessel in a fibroma. The calcified tissue is stained by haematoxylin (even after decalcification), and presents a dark granular appearance. $\times 500$.

(2) *Tissue death.* Calcification commonly occurs in (a) the necrotic lipid debris in atheromatous patches, (b) fat necrosis (usually around the pancreas or in the breast), (c) old infarcts (d) caseous patches in tuberculosis, necrotic foci in histoplasmosis and other chronic infections, (e) necrotic foci in malignant tumours, and (f) dead parasites (e.g. *Trichinella spiralis* and echinococcal cysts). Calcification of such dead tissue is a slow process, and

occurs only when necrotic material persists for a long time without undergoing organisation.

(3) *Inspissated pus and organic material in ducts, etc.* A large collection of pus, unless discharged, may eventually become inspissated, then calcified, and even ossified. Organic material accumulating in the ducts of salivary glands, or in the appendix, may become calcified, forming 'stones' in these sites. Calcium deposition in the urinary tract, both as discrete stones and as soft, crumbling material, is caused by urinary infections, but stone formation occurs also as a result of increased calcium excretion (see below).

(4) *Thrombi.* Calcification occurs very commonly in old venous thrombi which have not undergone organisation: hard masses are thus formed in veins, e.g. in the legs, and show up on x-ray as *phleboliths*.

The chemical reactions involved in dystrophic calcification are not understood. Factors which may be involved include the following. (a) Local changes in pH of hyaline or necrotic tissue, etc.: calcium is deposited more readily from an alkaline medium. (b) Breakdown products of cells or tissue elements to provide a nucleus with an affinity for calcium salts. Release of phosphate from nucleoprotein breakdown is a possible example. The strong tendency for calcification of necrotic fatty tissue was formerly explained by the affinity of fatty acids for calcium, forming insoluble calcium soaps. This suggestion lacks supporting evidence, and in particular subcutaneous injection of fatty acids does not lead to calcification. (c) Local enzyme changes: the normal process of calcification of growing bone occurs in the presence of high local concentrations of alkaline phosphatase. In experimentally induced lesions, some correlation has been observed between high levels of alkaline phosphatase and deposition of calcium salts, but the correlation is not a very good one, and this is not a convincing factor in dystrophic calcification in man.

Calcinosis circumscripta. This is a condition in which irregular nodular dystrophic calcification occurs in the skin and subcutaneous tissues, especially of the fingers. The overlying skin becomes ulcerated and the chalky material is discharged or may be scraped out. This appears to consist chiefly of calcium carbonate, as shown by solution with effervescence in hydrochloric acid. Microscopically a mild chronic inflammatory reaction with giant cells

surrounds the nodules. The causation of the lesion is obscure. The deposits are easily distinguished from gouty tophi by their dense opacity to x-rays and by histochemical tests. (See also tumoral calcinosis, p. 928).

Occasionally calcium deposition is more widespread, involving also muscles and tendons—this is known as **calcinosis universalis**.

A number of other diseases, including scleroderma and dermatomyositis, are occasionally complicated by calcification of the dermis or subcutaneous tissues.

Metastatic calcification

This occurs in the following conditions.

(1) **Excessive absorption of calcium from the gut**, seen most commonly in infants with hypervitaminosis D due to over-fortification of infant foods with vitamin D and calcium (p. 889). Similar experimental changes can be produced readily in the rat (Fig. 10.16).

Excessive intake can result also from taking very large amounts of calcium by mouth, for example milk and calcium carbonate by sufferers from peptic ulcer: this may lead to hypercalcaemia and alkalosis (the milk-alkali syndrome).

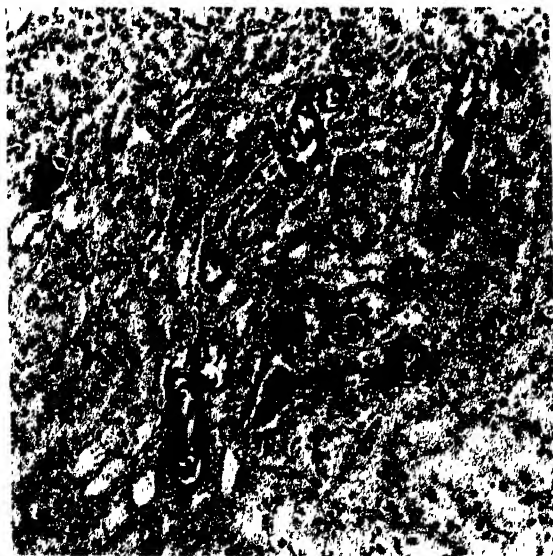


Fig. 10.16 Rat kidney in hypervitaminosis D. Note calcified small vessels and renal tubules. $\times 120$. (From preparation kindly lent by Dr. J. R. M. Innes.)

(2) **Excessive mobilisation of calcium from the bones**. This occurs in patients with *widespread bone destruction*, as for example in multiple myeloma or metastatic carcinoma. *Prolonged immobilisation* in bed for any reason is also of importance, the bones undergoing disuse atrophy. Excessive mobilisation of bone calcium is also brought about by *primary hyperparathyroidism*, usually due to a parathyroid adenoma (p. 1035), but a more common cause is *secondary hyperparathyroidism* associated with parathyroid hyperplasia and resulting from chronic renal failure with retention of phosphate (p. 1036).

Metastatic calcification occurs especially in the walls of arteries and in the kidneys. It is occasionally seen in the myocardium, acid-secreting gastric mucosa and the alveolar walls of the lungs. It may be that the sites of metastatic calcification are determined by a relatively high pH, e.g. around the renal tubules and the acid-secreting gastric glands. Deposition is seen initially on the surface of elastic fibres, basement membranes and other formed elements.

The **kidneys and urinary tract** are the commonest and most important sites of metastatic calcification. In the kidney, deposition of calcium occurs in the tubular epithelium and may be seen by electron microscopy in relation to mitochondria. More gross calcium deposition may occur in the tubular basement membranes, the interstitial tissue, and calcified concretions may form in the tubular lumen, usually of the collecting tubules. These changes are accompanied by impaired function, and by the development of coarse scars involving segments of the cortex and medulla (possibly due to obstruction of individual collecting tubules). Chronic renal failure may result. The renal arteries may also be calcified but, as elsewhere, patency is little affected.

Another important feature of hypercalcaemia is the formation of calcium carbonate/phosphate stones in the urinary tract (p. 868).

Nephrocalcinosis and stone formation are particularly liable to occur when there is increased intake of calcium associated with alkalosis and an alkaline urine, as in the milk-alkali syndrome; they occur also in renal tubular acidosis (p. 845).

Deposition of Uric Acid and Urates

Uric acid is formed as the final breakdown product of purine bases, and is thus derived from catabolism of nucleic acids. Normal plasma urate levels depend greatly on the assay technique, but levels above 7.0 mg/100 ml (0.42 mmol/l) for men and 6.0 mg/100 ml (0.36 mmol/l) for women are abnormally high. Adults produce 400–700 mg of endogenous uric acid daily and dietary purines contribute 300–600 mg. Most of this uric acid is excreted by the renal distal convoluted tubules, which can normally increase the rate of excretion, as necessary, to maintain homeostasis.

Hyperuricaemia is not uncommon, particularly in men over 40. It tends to be familial, but sporadic cases occur. The metabolic abnormalities concerned are not clearly understood. In some instances, increased production of uric acid results from a deficiency of the phosphoribosyl-transferase enzyme which is necessary for the re-utilisation of hypoxanthine for purine synthesis. This deficiency results in increased breakdown of hypoxanthine into uric acid. Other enzyme deficiencies with similar effect have been detected in some instances of hyperuricaemia. In others, there is a defect of unknown nature in renal excretion of uric acid. These defects account for at least some cases of *primary hyperuricaemia* in which nucleic acid breakdown is normal. *Secondary hyperuricaemia* results from increased nucleic acid breakdown, as in chronic myeloid leukaemia (p. 547).

There is considerable variation in the effects of hyperuricaemia. In most instances, there are no associated pathological changes. In others there is deposition of uric acid or urate in the collecting tubules of the kidneys, seen macro-

scopically as brown-yellow streaking of the medulla: this may have little or no effect, or may be followed by formation of uric acid stones (p. 868). Uric acid streaking of the medulla is a common necropsy finding, particularly in children, and appears to be associated with a state of dehydration before death. The most important complication of hyperuricaemia is **gout** (p. 923), in which crystals of monosodium urate are deposited in and around the joints, in the skin (Fig. 10.17) and elsewhere. It is always accompanied by hyperuricaemia, and yet the relatives of patients may have equally high levels of plasma uric acid without developing gout. As indicated above for hyperuricaemia in general, a number of individual abnormalities of purine metabolism can result in gout.



Fig. 10.17 Section through gouty nodule of skin, showing deposit of needle-like crystals of monosodium urate. $\times 370$.

Further Reading

- Bothwell, T. H., Charlton, R. W., Cook, J. W. and Finch, C. A. (1979). *Iron Metabolism in Man*, pp. 576. Blackwell Scientific, Oxford, London, Edinburgh and Melbourne. (A comprehensive review by four leading experts, with an extensive bibliography.)
- Glennier, G. G. and Page, D. L. (1976). Amyloid, Amyloidosis and Amyloidogenesis. In *International Review of Experimental Pathology*, Vol. 15, pp. 1–92.
- Lendrum, A. C. (1969). The Validation of Fibrin and its Significance in the Story of Hyalin. In *Trends in Clinical Pathology*, pp. 159–183. British Medical Association, London.
- Scheinberg, M. A. and Cathcart, E. S. (1976). Comprehensive study of humoral and cellular immune abnormalities in 26 patients with systemic amyloidosis. *Arthritis and Rheumatism*, **19**, 173–182.

Tumours: 1. General Features, Causation and Host Reactions

General Features of Tumours

In previous chapters we have seen examples of cell proliferation and growth of tissues in the process of repair, in response to irritation, and as a hyperplastic response to increased workload or hormonal stimulation. Such growth is purposeful, and, up to a point, capable of explanation. In a tumour (neoplasm), however, the growth is not only excessive but apparently purposeless, progressing without regard to the surrounding tissues or the requirements of the individual as a whole. While forming a part of the body, tumour cells seem to have become largely unresponsive to the factors which control the proliferation of non-neoplastic cells. Accordingly, tumours exhibit various degrees of uncontrolled growth and in some instances uncontrolled function, e.g. the production of hormones or enzymes. Such behaviour is commonly termed *autonomous*, but a tumour is, of course, dependent on the host for its nutrition, blood supply and supporting stroma, and escape from host control factors is only relative.

Definition. A tumour, or neoplasm, is an abnormal mass of tissue, the growth of which exceeds and is unco-ordinated with that of the normal tissues and continues in the same manner after cessation of the stimuli which have initiated it. This definition covers most tumours, which form discrete lumps, but in the leukaemias, which are tumours of myeloid or lymphoid cells, the tumour cells may extend diffusely through the marrow or lymphoid tissues, and also circulate in the blood.

Origin. Tumours show an extraordinary variety of structure, but the majority retain a resemblance to some normal tissue or cell type;

occasionally the resemblance is to some precursor cell or tissue rather than to the fully differentiated adult type. These resemblances are attributable to the origin of each tumour from abnormal and excessive proliferation of a cell derived from the previously normal tissue. Tumours arise most often from tissues in which the cells are normally labile (i.e. they are continually being replaced by new cells—p. 77), and which are exposed to the various noxious agents in the environment (especially the skin and the epithelium of the alimentary and respiratory tracts). Many tumours do, however, originate from the cells of organs not so exposed and normally having more stable cells, e.g. those of the liver, thyroid, adrenal, cartilage or fat. The adult neuron is probably the only type of nucleated cell in the body incapable of giving rise to a tumour.

Classification. The cell or tissue origin of a tumour is called its *histogenesis*, and provides the basis of a principal mode of classification. On this basis, nearly all tumours may be classified as *epithelial* or *connective tissue* tumours according to the cell of origin. Tumours are further classified by their naked-eye appearances, their microscopic features and by the nature of their products. *The most important mode of classification is, however, based on behaviour, and divides tumours into benign and malignant types (see below).* So little is known of the precise causal factors of most individual tumours that a classification based on aetiology is not yet widely applicable.

In this text, **tumour** or **neoplasm** is used for all lesions of this type, whether benign or malignant. **Cancer** is used for all malignant tumours, regardless of their origin. **Carcinoma** is used only for malignant tumours of epith-

elium, and **carcinogenesis** for the changes involved in the development of malignant tumours of all types.

Variations in tumour behaviour

Tumours vary considerably in their behaviour, notably in the following important features.

Rate of growth. There are all gradations between slowly-growing tumours that hardly change in size from year to year and those which grow so rapidly that differences in size may be detected from week to week. Many tumours consist very largely of tumour cells, with blood vessels and supporting stroma contributing little to the total mass: their rate of growth depends on the rate of proliferation and the life-span of the tumour cells. Epithelial tumours may add to their bulk by accumulation of material secreted by the tumour cells, e.g. mucin, or they may induce a fibrous reaction, so that they come to consist largely of fibrous stroma in which lie groups of tumour cells. The growth of some connective tissue tumours depends largely on the production of matrix (collagenous, cartilagenous, etc.) by the tumour cells. Vascular congestion, oedema and infection can all occur in tumours, as in normal tissues, and contribute to fluctuations in their rate of growth.

By definition, the *rate of cell production* in tumours exceeds the rate of cell death. The rate of cell production depends on the *growth fraction*, i.e. the proportion of cells entering the cell cycle which culminates in mitosis (Fig. 11.1), and the time taken to complete the cycle. The number of mitoses seen microscopically is thus a general indication of the rate of tumour growth. In general, the rate of cell production is greater in malignant than in benign tumours. The kinetics of tumour cell proliferation are of importance in planning certain types of therapy, for many anti-tumour drugs destroy mainly cells undergoing mitosis. The features determining the *life-span of tumour cells* are complex and not fully understood. In some tumours, for example basal cell carcinoma of the epidermis, many cells undergo shrinkage necrosis (*apoptosis*—p. 11), and in spite of a high mitotic rate, growth is surprisingly slow. In many malignant tumours *disorders of mitosis* result in abnormalities in the number and

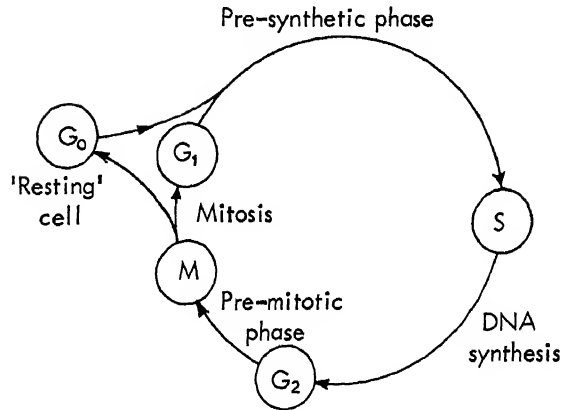


Fig. 11.1 The mitotic or cell cycle, showing the relative duration of each phase. The whole cycle normally takes from about 24 hours to several days, depending largely on the duration of G_1 , the phase preceding the synthesis of DNA (S phase). The S phase is followed by a brief (G_2) phase before the cell undergoes mitosis (M). After mitosis, the cell may leave the cycle to enter the 'resting' (G_0) stage, in which it may remain for its whole lifespan; cells in G_0 phase may, however, re-enter the cycle at the G_1 phase. The factors which inhibit cell division do so by arresting cells in the G_0 or G_1 phase.

structure of the chromosomes. The cells produced are pleomorphic (Fig. 11.3) and many of them are non-viable. *Aberrant mitoses* are thus a feature of malignant tumours. Malignant tumours also tend to outgrow their blood supply and the rapidly increasing number of cells compress the small blood vessels. Accordingly, *ischaemic necrosis* is a conspicuous feature of many malignant tumours (Fig. 11.2); those involving the skin or a mucous membrane tend to ulcerate and *bacterial infection* then results in more extensive necrosis.

Although many malignant tumours grow progressively and relentlessly, individual tumours may fluctuate greatly in their growth rate. It is, for example, not uncommon for removal of a breast cancer or a melanoma of the skin to be followed by many years of good health: in some patients, local or distant foci of residual tumour eventually become apparent and grow rapidly. Clearly, tumour has persisted since before the time of excision of the original tumour, but for many years has failed to grow significantly. This suggests that host defence mechanisms are involved, and that they may arrest tumour growth for long periods.

Invasion and spread. *The cells of benign tumours remain at the site of origin, forming a single*

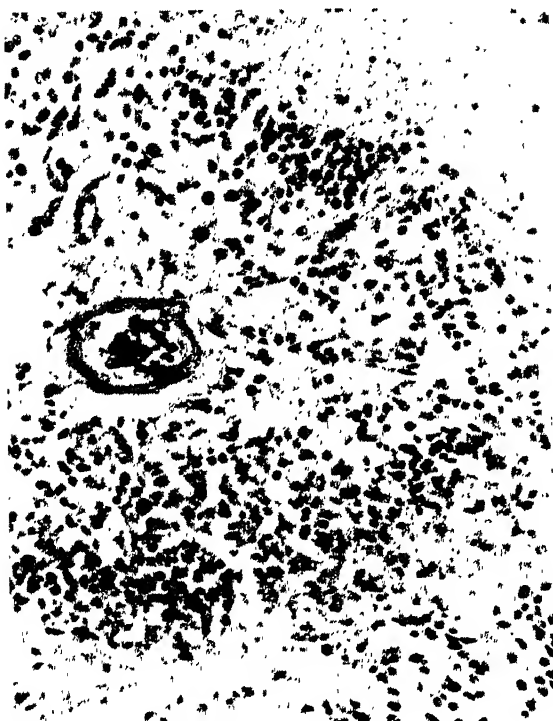


Fig. 11.2 Ischaemic necrosis in a tumour. In this example, the tumour is a rapidly growing anaplastic astrocytoma of the brain, composed of cells with a round dark nucleus. The tumour cells immediately around a small blood vessel have survived, forming a cuff which occupies most of the picture. More peripherally the tumour has undergone ischaemic necrosis with loss of nuclear staining. $\times 190$. (Dr. A. M. Lutfy.)

mass. When growing in a solid tissue, they compress the surrounding normal cells which undergo pressure atrophy and necrosis: the tissue stroma is more resistant and may become condensed to form a fibrous *capsule* (Fig. 12.8, p. 326). The formation of a capsule has however been given too much emphasis, for in some benign tumours the capsule is incomplete, little or no stroma separating the adjacent tissue from the tumour cell mass: this is particularly the case in tissues like the adrenal which can expand readily without pressure atrophy occurring.

The cells of malignant tumours invade locally and also spread by the lymphatics, bloodstream and body cavities to form secondary tumours or metastases remote from the site of origin. In some instances it is difficult or impossible to determine which is the original or **primary tumour**, particularly when their cells are poorly differentiated (see below).

Differentiation. This is the degree of resemblance of a tumour to its tissue of origin and can be applied morphologically and functionally to the tumour cells. The naked eye appearances of a well differentiated tumour sometimes reveal its nature: for instance a lipoma is usually recognisable as an encapsulated mass of adipose tissue of essentially normal microscopic appearance. The less differentiated a tumour is, the more difficult it is to identify its tissue of origin and the tissue of origin of a poorly differentiated tumour may remain unknown despite naked eye, light- and electron-microscopic and biochemical examination. Partial loss of differentiation is termed **dysplasia** and complete loss, so that the tumour no longer resembles its tissue of origin, is termed **anaplasia** (Fig. 11.3).

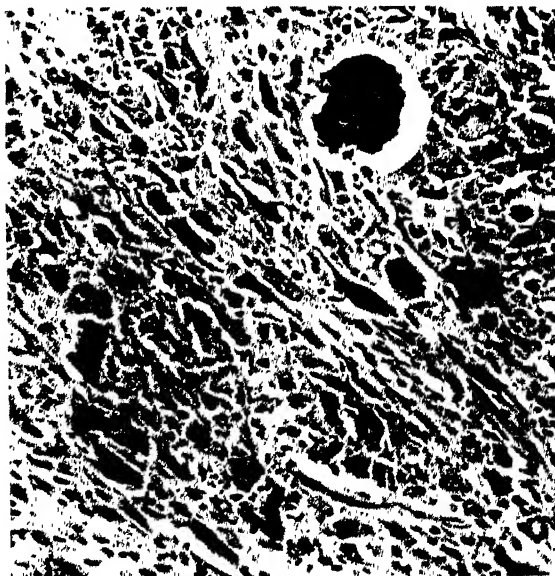


Fig. 11.3 Section of an anaplastic malignant tumour showing also great variation in shape and size of the cells and of their nuclei (pleomorphism). Note also the abnormal mitoses. $\times 200$.

There is a general correlation between the above features; *malignant tumours are usually rapidly growing, poorly differentiated, and have a high mitotic rate with nuclear pleomorphism and abnormal mitoses* (Fig. 11.3), while *benign tumours are usually slow growing, well differentiated, and show infrequent mitoses and little cytological variation* (Fig. 11.4, see also Figs. 12.2–20, pp. 323–32 and 13.2–15, pp. 341–7).

Benign tumours seldom kill unless they arise near and press on vital structures or secrete ex-

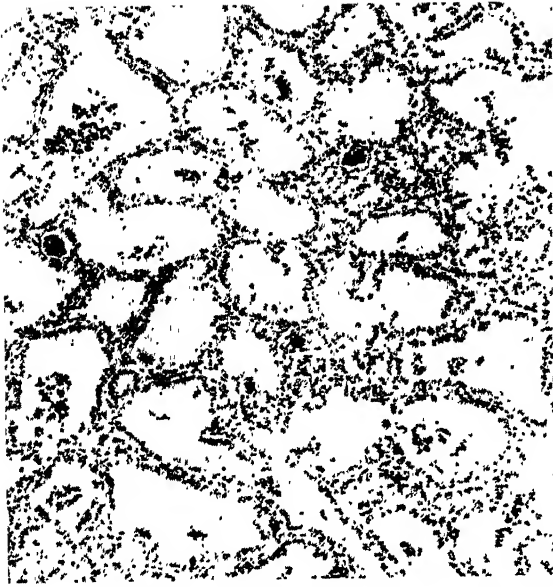


Fig. 11.4 Section of a benign tumour of the thyroid. Note the close resemblance to thyroid tissue. $\times 75$.

cessive amounts of hormone. Most fatal tumours are malignant, and death may result from local invasion, from the effects of metastases, or from a combination of both. The identification of tumours as benign or malignant is therefore crucial and is a major responsibility of the hospital histopathologist. Table 11.1 compares the characteristics of benign and malignant tumours. However, the distinction is not absolute and borderline tumours occur. Some tumours, for example, are locally invasive and yet rarely metastasise.

Progression of a benign tumour to malignancy is not common, and most malignant tumours do not arise from a benign tumour. The behaviour of the innumerable different kinds of tumours of different organs varies greatly, and it is necessary to know these variations before one can apply the criteria in Table 11.1 with safety to the individual patient.

Examples of exceptional behaviour. Rate of growth. Some benign tumours (especially of the female genitalia, e.g. myomas of the uterus and cystadenomas of the ovary) may grow very rapidly and reach a great size. Some malignant tumours—basal cell carcinomas of the epidermis, some breast carcinomas and in particular carcinoid tumours of the gut and 'latent' carcinoma of the prostate—grow very slowly. Incidentally, the normal fetus grows faster than any tumour.

Table 11.1 Contrasting features of benign and malignant tumours

	Benign	Malignant
(a) <i>Evidence on rate of growth</i>		
Mitoses	Few and normal	Numerous and often abnormal
Nuclei	Little altered	Enlarged, often irregular (pleomorphic)
Nucleoli	Little altered	Usually large
Cytoplasmic basophilia	Slight	Marked
Haemorrhage and necrosis	Inconspicuous	Often extensive
(b) <i>Differentiation</i>		
Naked-eye resemblance to tissue of origin	Often close	Variable: from close to none
Microscopic resemblance to tissue of origin	Usually very marked	Usually poor
Function, e.g. secretion	Often well maintained	May be retained, lost, or abnormal products
(c) <i>Evidence on transgression of normal boundaries</i>		
Capsule intact	Frequent	Rare (usually none)
Local invasion	Absent	Very frequent
Metastases	Never	Frequent

Differentiation. Some benign tumours show a high degree of cellular specialisation and structural arrangement, but do not resemble the parent tissue. Mucinous cystadenoma of the ovary (Fig. 24.21, p. 962) is an example: it is 'well differentiated' but in a different direction from the parent tissue. Some malignant tumours (some squamous carcinomas, for instance, and well differentiated thyroid carcinomas) may closely resemble the parent tissue. Function is not always lost in malignant tumours; indeed, such normal functions as the production of keratin, mucus, melanin, and hormones, may be well maintained by malignant tumours.

Invasion is perhaps the most nearly reliable criterion of malignancy, but it is often surprisingly difficult to assess in practice, especially where the normal structures are distorted by some other pathological process such as infection, metaplasia or congenital anomaly. As already mentioned, a capsule is absent in some benign tumours, and some malignant tumours are encapsulated, including clear-cell carcinoma of kidney and some thyroid carcinomas.

Metastasis is another generally reliable criterion

ion, but benign tumours and even normal tissue may sometimes become implanted at a distance as a result of trauma or surgical accident. The placental trophoblast not only invades the uterus but is often carried by the blood to the lungs. Also some undoubted malignant tumours practically never metastasise: basal cell carcinoma is the best example, but intracranial tumours also fail to metastasise outside the cranio-spinal cavity.

Despite these exceptions, the hallmark of the malignant tumour is its capacity to spread to, and grow progressively in, tissue remote from its site of origin. Spread may occur by lymphatic vessels as tumour cell emboli, or a column of tumour cells may grow along the vessel until a lymph node is reached. Invasion of a lymph node may be followed by further spread via the efferent lymphatic. Similar invasion and spread by blood vessels is also common. Tumour cells disseminated by the bloodstream may involve any organ, but the lungs, liver and bone marrow are specially common sites of secondary tumours. (It should be noted, however, that by no means all tumour cells which enter the bloodstream go on to establish metastases. Circulating tumour cells can be detected in the bloodstream of patients with early cancer and it seems that many such cells are destroyed). Less common but important routes of spread are across body cavities (transcoelomic spread) and intra-epithelial extension as in Paget's disease of the nipple. The spread of tumours is discussed more fully in Chapter 12.

Effect of tumours

These are various and many of them can be readily understood.

Local effects. The presence of a mass of growing tissue of whatever kind may lead to **pressure effects** on various important structures, e.g. on blood vessels (especially veins), nerves, hollow viscera and ducts and solid organs, resulting in a wide variety of complications. This is true both of benign and of malignant tumours, but in addition the latter infiltrate and destroy such structures, and are especially liable to produce obstructive effects, e.g. stenosis of pylorus, intestine or bronchi.

As mentioned earlier, extensive necrosis commonly occurs in malignant tumours: those involving the skin or mucous membranes very often ulcerate and become infected by bacteria

to which they are more susceptible than normal tissues.

Widespread replacement of organs by tumour tissue may impair their function: involvement of bones leads to fractures: direct invasion or compression of nerves causes much of the pain associated with malignant disease. Compression or infiltration of blood vessels or lymphatics leads to regional congestion, ischaemia and oedema.

General effects. Absorption of bacterial products from infected tumours and of the products of necrosis of tumour tissue contributes to the pyrexia, debility and wasting (**cachexia**) seen in some cancer patients. Where there is a large volume of tumour, a further factor in producing cachexia is the competition between tumour and normal tissues for essential nutrients, such as amino acids and vitamins. Actual reduction of food intake is important in patients nauseated from liver metastases or the effects of cytotoxic drugs, or with dysphagia resulting from neoplastic involvement of the upper alimentary tract. There is little direct evidence that tumours commonly produce toxic compounds, although some patients with cancer develop unusual types of neuropathy, etc. (see below), which are likely to be due to abnormal tumour products.

Anaemia is common in cancer patients; it can result from haemorrhage from, or infection of, an ulcerated tumour, replacement of the haemopoietic marrow by tumour, or marrow depression by cytotoxic drugs or radiotherapy.

Malignant tumours cause **depression of immunological and other defence mechanisms**. This depression, which may be increased by chemotherapy or radiation, predisposes patients with cancer to **infection** with both virulent and opportunistic pathogens.

Occasional effects of cancer include various ill-defined **neuropathies and myopathies** which interfere respectively with the functioning of the nervous system and skeletal muscles: they are most likely due to humoral products of tumours. Other remote effects include **multiple venous thromboses** (especially in pancreatic cancer) and **various skin rashes**. **Renal disturbances** (usually nephrotic syndrome) occasionally result from deposition of tumour-antigen/antibody complexes in the glomeruli. These and other effects of tumours will be exemplified in the accounts of individual systems and organs.

Hormonal effects. *Syndromes of hormone excess* may result from the production of large quantities of hormone by benign (and less commonly malignant) tumours of the endocrine organs. These effects may be regarded as appropriate as they reflect 'appropriate' functional differentiation of the tumour cells. Much more surprising is the increasing list of hormones shown to be produced by some tumours arising from tissues with no known relevant hormone secretion. The most commonly encountered examples are production of hormones with ACTH or ADH activity by carcinomas of the bronchi. Syndromes due to such '**inappropriate**' secretion of hormones by tumours of apparently non-endocrine origin occur in relatively few patients with cancer, but a much higher proportion of tumours can be shown to have the enzyme systems necessary for the production of such hormones. The tumours most commonly associated with the inappropriate

secretion of hormones are listed in Table 11.2.

The basis of the 'inappropriate' secretion of hormones by non-endocrine tumours remains unclear. Normal somatic cells contain the whole genome of the individual, and during differentiation the genes not required by each particular cell type are suppressed. Apparently the nuclear changes in tumour cells sometimes include re-expression of suppressed genes, but the association between particular types of tumour and hormone production cannot readily be explained by random de-repression. A recent and attractive theory holds that there are widely distributed cells with an endocrine function ('apud' cells) and that tumours arising from such cells ('apudomas') amplify and make detectable their actual or potential hormonal activities (p. 1034).

By invading and destroying endocrine glands, tumours can also cause hormonal deficiencies.

Table 11.2 Examples of 'inappropriate' hormone secretion by tumours

Hormone secreted by tumour	Type of tumour
ACTH	Oat cell carcinoma of bronchus; epithelial thymomas; carcinoid tumours; islet cell tumour of pancreas.
Parathormone	Squamous carcinoma of bronchus; carcinomas of oesophagus, colon, liver, pancreas, kidney.
Antidiuretic hormone (ADH)	Oat cell carcinoma of bronchus; haemangioblastoma of cerebellum.
Insulin	Retroperitoneal fibrosarcoma; mesothelioma; hepatoma; adrenal carcinoma.
Thyroid stimulating hormone	Choriocarcinoma; hydatidiform mole; embryonal carcinoma of testis.
Erythropoietin (erythrocytosis)	Renal carcinoma; cerebellar haemangioblastoma; hepatoma; pheochromocytoma.
Gonadotrophin (precocious puberty in males)	Hepatoma.

The Causation of Tumours

Paradoxically we know many causes for cancer, but not *the* cause of cancer. We can detect many changes in the cancer cell, but we do not know the nature of the essential change in the cell which makes it a cancer cell. We know that various chemical compounds, x-irradiation and (in animals) viruses can produce tumours but we do not know exactly how any of them renders cells neoplastic.

The process of conversion of a normal cell to malignancy is called **carcinogenesis** and agents

which cause this are termed **carcinogens**. Carcinogenesis in man is nearly always a complex process, usually involving the interaction of many factors, some of which favour tumour development and others which appear to provide some protection against it. They may be divided into (1) **genetically determined factors**, which in total determine an individual's susceptibility to develop a particular cancer on exposure to (2) the **exogenous influences** encountered in the complex environment in which

we live. The complexities of genetic and environmental factors and of their interactions account for many of the difficulties of the epidemiological and experimental investigation of the causes of cancer.

There is great variation in the intensity and length of exposure to individual carcinogens necessary to bring about tumour development. A subthreshold dose of a carcinogen will not produce a tumour, but subthreshold doses of two separate carcinogens given together may be effective (**syncarcinogenesis**). The combination of certain substances which are not of themselves carcinogenic (**co-carcinogens** or **promoters**) with a subthreshold dose of a carcinogen will also cause tumour development (**co-car-**

cinogenesis) For example, if a chemical carcinogen such as methylcholanthrene is painted on the skin of a mouse, application of a dilute solution of croton oil (itself not carcinogenic) will hasten the development of tumours and increase the number which develop—only, however, if applied together with or after the carcinogen. The carcinogen thus appears to *initiate* an irreversible process, while the co-carcinogen (in this case croton oil) *promotes* its progress after initiation. In most experimental studies, administration of a sub-threshold dose of carcinogen and a co-carcinogen has resulted in benign tumours and the role of co-carcinogens in human cancer is uncertain.

Genetic factors

The share of genetic factors in cancer causation varies extremely widely, from almost negligible in some common cancers to practically 100% in a few rare tumours, with many intermediate positions.

Clearly there must be strong selective pressure against genes producing major cancers in childhood and young adults, and such conditions are always rare. *Polyposis coli* (p. 651) is a good example of the tumours which escape this pressure. It is determined by a Mendelian dominant factor, multiple polyps developing in the colon in half the members of affected families; cancer of the colon develops regularly in early adult life—late enough, however, to permit reproduction. No known environmental factor is involved, though it is possible that the unknown basic abnormality of the colonic epithelium involves genetically-determined susceptibility to some material present in the bowel contents. *Retinoblastoma* (p. 803) is also determined by a Mendelian dominant factor, and so are the multiple benign tumours (only occasionally becoming malignant) of *neurofibromatosis* (p. 794), and also *Peutz-Jegher's syndrome* (p. 650) in which benign tumour-like polyps develop, usually in the small intestine.

A particularly illuminating example of the interaction of a genetic factor and environment is *xeroderma pigmentosum*. Sufferers show severe sunburn on minimal exposure to sunlight, and develop multiple skin cancers, ultim-

ately fatal, often while still in their teens. The condition has been shown to be due to a simple enzyme defect, the absence of an endonuclease that removes abnormally linked pairs of bases in the DNA chain and replaces them by normal pairs. Such linkages are produced in *normal skin* in enormous numbers by exposure to the UV light in ordinary sunlight, and require constant repair by the endonuclease. In the absence of the normal repair mechanism, many cells die and others undergo permanent genetic damage that ultimately leads to tumour growth in some of them. It can therefore be said that the cancer of xeroderma is in fact caused by the external carcinogen, UV light, the genetic defect being simply the absence of a defence mechanism against this agent. This is confirmed by the great improvement in life expectancy produced in these cases by rigorous protection from sunlight.

The relationship of *skin cancer* to skin pigmentation is also instructive. The defence mechanism just mentioned fails in normal skins if exposure to UV light is maintained at high levels for a lifetime. Races exposed to much sunlight have developed more or less heavy melanin pigmentation as a defence (its function being confirmed by its localisation as caps over the epidermal-cell nuclei). The white races evolved in high latitudes where there is much less exposure to UV, and so little danger of skin cancer. Rickets is a powerful factor select-

ing against dark-skinned races in high latitudes because it causes distortion of the female pelvis and so endangers mother and infant during childbirth. Recent emigrations demonstrate these factors: celtic types in Australia have a colossal incidence of basal cell carcinomas of the skin (present in 75% at the age of 75 in some areas), while in Glasgow some Pakistani children develop rickets unless their diet is supplemented by vitamin D. We have here another example of a 'normal' genetic variant which greatly influences the production of cancer by the external carcinogen, UV light.

The interspecies differences in enzyme handling of 2-naphthylamine, which influence the incidence of the urinary bladder cancer (p. 299) fall into this same category of gene-environment interaction in cancer production. Recent work suggests a similar factor in lung cancer in man. The carcinogenicity of cigarette tar hydrocarbons depends on their conversion to epoxides by an enzyme, aryl hydrocarbon hydroxylase (AHH), the concentration of which varies considerably in different individuals; some studies suggest that the incidence of lung cancer is highest in those with high AHH levels.

A more remote connection is shown by the relation of some tumours to *genetic markers*—gastric cancer with blood group A, Hodgkin's disease with the HLA antigen B5 and acute lymphoblastic leukaemia with HLA antigens A2 and B12: all that is known with certainty is the raised frequency of the tumour in bearers of the gene concerned. In all three examples the effect is small and the mechanism obscure.

A high incidence of a particular cancer, e.g. of the breast, stomach or colon, has been observed in some families, but the great majority of such cancers arise in members of families without any such predisposition. The relative importance of genetic and environmental causal factors in cancer-prone families is not known.

Chromosomal abnormalities in tumour cells

The abnormal behaviour and morphology of tumour cells seems likely to result from alterations in the numbers, arrangement or operation of genes. Morphologically, we are limited at present to seeking evidence of abnormalities in chromosomal morphology and number: we

cannot yet detect morphological changes in individual genes, even with modern chromosomal banding techniques (Fig. 2.8, p. 16).

Chromosomal abnormalities cannot be detected in most benign tumours, but in malignant tumours the number of chromosomes is often abnormal and structural chromosomal abnormalities are found. However, these changes seem to occur in random manner. As they are more marked in advanced than in early tumours, they probably merely reflect the progressively disordered nuclear behaviour consequent upon the development and evolution of cancer. Of more interest are chromosomal abnormalities which occur in some benign and malignant tumours in a non-random fashion. The outstanding example is the Philadelphia chromosomal abnormality, involving C22 and C9 (p. 549), which is demonstrable in 90% of cases of chronic myeloid leukaemia (CML). The change affects haemopoietic stem cells and is now believed to be the basic cell change leading to this form of CML. It is also of interest that many cases of CML pass into a more acute, malignant form of leukaemia, and in about 75% of such cases the acute phase is accompanied by further non-random chromosomal changes, usually involving C8 and C17. CML without the Philadelphia chromosome abnormality may well be a separate entity, for the course and response to treatment are different. Other examples of non-random chromosomal changes associated with particular forms of neoplasia are known, but apart from CML, they appear likely to develop as secondary events, for they are often absent from the tumour cells obtained by early biopsy.

Interesting observations have been made on the effects of cell hybridisation (p. 304) on malignant cells. When cancer cells are fused with normal cells, the hybrid cells are often non-malignant, as assessed by their failure to produce tumours in suitable animals or to grow progressively with loss of contact inhibition in culture (p. 302). The hybrid cell is, however, unstable, and in culture it loses chromosomes with successive divisions. Loss of particular chromosomes derived originally from the normal cell has been found to be associated with the re-emergence of malignant behaviour. It thus appears that the factors which make a cell malignant may be present but suppressed by genetically-determined factors of the normal

cell: they may become operative with loss of the suppressor factors.

The development of malignancy as a result of chromosomal aberrations is suggested also by

the increased incidence of cancer in rare congenital disorders characterised by chromosomal instability, e.g. Fanconi's anaemia, ataxia telangiectasia (p. 171) and Bloom's syndrome.

Chemical carcinogens

In 1775, Percivall Pott*, a London surgeon, recorded a high incidence of cancer of the scrotum in chimney sweeps (Fig. 11.5). Because they lacked facilities (and perhaps enthusiasm) for washing, sweeps retained soot in their rugose scrotal skin, where it exerted a carcinogenic effect. Another occupational cancer was reported in 1874 by Volkmann, who observed a high incidence of skin cancer in workers exposed to tar and mineral oil. Numerous other examples of chemically-related occupational cancers have since been reported.

Testing for chemical carcinogenicity. A major advance in the investigation of chemical carcinogenesis was made by Yamagiwa and Ichikawa who in 1917 reported the induction of cancers by repeated painting of rabbits' ears

with tar. This provided the means to test individual chemicals for carcinogenicity, and since that time a very large number of compounds have been tested. The basic experimental design has varied little. Suspected carcinogens are administered at regular intervals, (e.g. daily, thrice weekly, etc.) by whatever route seems most appropriate, either alone or together with co-carcinogens, and with or without manipulation of the animals' hormonal, metabolic or immunological status.

A large number of chemicals has been shown to be carcinogenic, and the following principles have been established.

1. There is a latent period between the first administration of the carcinogen and the development of tumours. The latent period varies

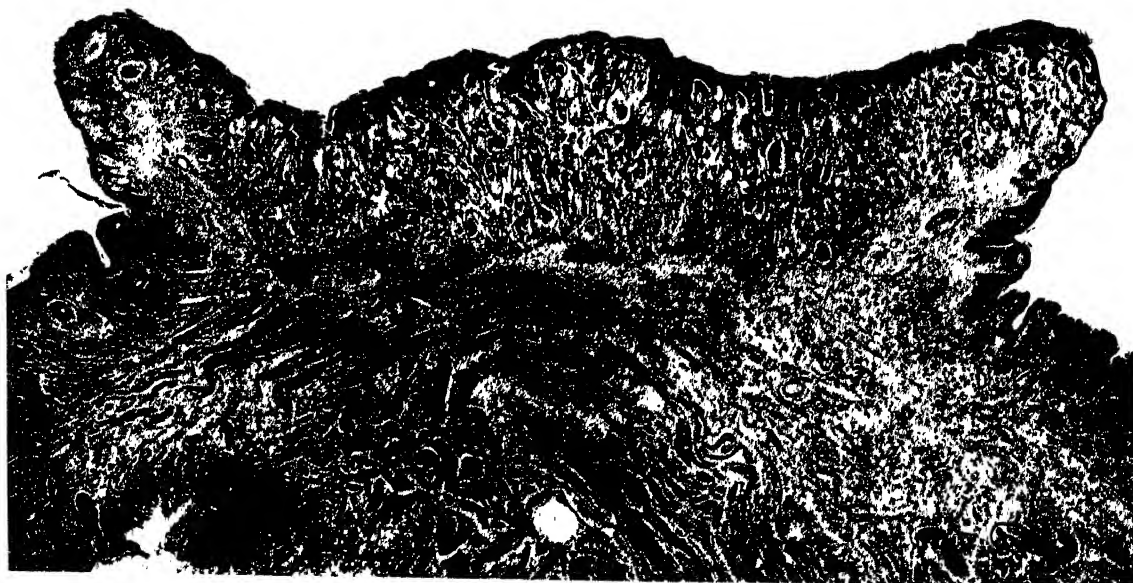


Fig. 11.5 Cancer arising from the epidermis of the scrotum in a chimney sweep. The tumour forms a raised plaque. Numerous small groups of tumour cells are seen invading the dermis. The dark fibres in the deep part of the tissue are cremasteric muscle. $\times 5$.

* Percivall Pott's name is also commemorated by Pott's fracture and Pott's disease of the spine.

with different carcinogens and with the same carcinogen for different species and strains of animal.

2. For some carcinogens, the latent period varies inversely with the size of each dose of carcinogen over a wide range of dosage. Doubling the dose will thus half the time taken for tumours to appear. This means that the effect of each dose is irreversible and that tumours appear after a certain *total* dose of the carcinogen. There is thus a *threshold dose* for each carcinogen/strain combination.

3. For other carcinogens, the latent period is influenced by the dosage, but to a smaller degree, so that doubling the dosage does not reduce the latent period by half. In these instances, very small doses will induce tumours if the animals live long enough, and it has been concluded that, for such substances, there is no sub-threshold dose, but merely a lengthening of the latent period.

4. Following administration of the threshold dose of carcinogen, there may in some instances be no morphological changes in the target cells, but tumours will eventually develop, even without further administration of carcinogen.

5. Some carcinogens act specifically on one organ, e.g. the liver; this is usually because the carcinogen is concentrated in that organ, or because it is really a *procarcinogen* which is converted to a carcinogen in that particular organ. Other carcinogens are non-organ-specific, and induce tumours at the sites of maximal exposure, which are usually determined by the route of administration.

6. The physical properties and chemical reactivity of the carcinogen are important, for they determine whether it diffuses throughout the body or stays at the site of administration, and also whether it is concentrated in certain organs and in the urine.

Testing chemicals for carcinogenicity is a complex, exacting procedure and, as illustrated below, it may take years to establish that a substance is carcinogenic. Carcinogenicity also varies considerably with species. An example is provided by the induction of cancer of the urinary bladder by 2-naphthylamine. This substance came under suspicion because a high incidence of bladder cancer was observed in workers in the aniline dye industry and analysis of the jobs of those workers developing

cancer pointed to 2-naphthylamine as the causal agent. Tests in laboratory animals of several species failed to demonstrate its carcinogenicity, but eventually it was shown to induce tumours of the bladder in dogs after *five years* oral administration. It was subsequently shown that dogs and men convert the bulk of the non-carcinogenic 2-naphthylamine to carcinogenic 1-hydroxy-2-naphthylamine in the kidney. This compound is concentrated in the urine and can induce carcinoma in the transitional epithelium lining the urinary tract, and particularly, because of its relatively large surface area, in the bladder. Most species of animals convert 2-naphthylamine to non-carcinogenic substances such as glucuronides, and this explains the initially negative studies of 2-naphthylamine carcinogenicity.

Testing for mutagenic activity. As exemplified above, carcinogenicity tests often take a long time, and even when very thorough they cannot provide an absolute guarantee that a compound is non-carcinogenic for man. More recently, tests have been devised to determine whether chemical compounds are capable of causing mutations in bacteria. By selecting suitable bacterial strains, tests for mutations of various types can be made, both with a suspected compound and with its metabolites. Such tests are rapid, and for most of the substances which have been tested there is a very good correlation between carcinogenicity and mutagenicity.

The nature of chemical carcinogens

Thousands of chemicals can induce the development of tumours. The first to be proved carcinogenic was the polycyclic hydrocarbon, 1:2, 5:6 dibenzanthracene, which was synthesised by Kennaway and his colleagues approximately 50 years ago. The first carcinogen to be identified in soot and tar was 3:4 benzpyrene. Since that time, many other **polycyclic hydrocarbons** have been proved to be carcinogenic. These compounds produce cancer at the site of local application. They are present in mineral oils and are released by combustion of most organic materials, including fossil fuels; accordingly, they are widespread atmospheric pollutants.

Other classes of compounds which are car-

cinogenic include **aromatic amines** used in the aniline dye industry, e.g. 2-naphthylamine (see above) and benzidine (formerly used also in the laboratory to detect blood in the faeces), which cause cancer of the urinary bladder. **Azo-dyes**, e.g. 'butter yellow' formerly used to tint margarine, cause liver cancer (see below), and **aminofluorenes**, notably 2-acetylaminofluorene, can induce cancer of the liver and of the urinary bladder. In each of these classes of compound, *carcinogenicity cannot usually be predicted from the molecular structure, and small changes in the molecule can greatly influence carcinogenicity*. It is, however, apparent that certain molecular features, for example double-bonds at certain sites in the benzene rings of the polycyclic hydrocarbons, are of importance.

Certain non-aromatic organic compounds are also carcinogenic. They include some **alkylating agents**, e.g. mustard gas and methyl-nitrosourea, and **nitrosamines** which may be formed from chemical action on foodstuffs in the stomach and bacterial action in the large intestine.

Various inorganic compounds also act as carcinogens; for example inhaled **asbestos** fibres cause cancer of the pleura, peritoneum and bronchus, while cancer of the bronchus can apparently be caused also by inhalation of dusts containing compounds of **arsenic**, **beryllium**, **nickel** and **chromates**. Long-continued administration of arsenical compounds or exposure to arsenical dusts can also cause skin cancers.

The mechanism of chemical carcinogenesis

This is poorly understood. All carcinogens do appear, however, to induce progressive and irreversible changes in the target cells, leading eventually to the development of a cancer cell. This is illustrated by studies of the effects of butter-yellow (p-dimethylaminoazobenzene), the most carcinogenic of the azo-dyes, on the liver. The dye is taken up by liver cells and, like all carcinogens, it causes cell injury resulting in this case in loss of some liver cells and compensatory regeneration of surviving cells. There is thus cellular proliferation which, as mentioned earlier, is important in carcinogenesis. If slices of normal rat liver are placed in culture

medium and oxygen uptake is measured as an index of cellular metabolism, addition of butter-yellow to the medium is found to depress metabolism. Administration of butter-yellow to rats progressively diminishes the effect of butter-yellow on the uptake of oxygen by liver slices *in vitro*, and when liver tumours finally develop the oxygen consumption is found to be unaffected by butter-yellow. It may thus be concluded that the liver cells become progressively adapted to the presence of butter-yellow *in vivo*. Such adaptation could be explained either (a) by the development of increased amounts of enzymes which detoxify butter-yellow or provide metabolic pathways not affected by it, or (b) by the production of mutations by butter-yellow and growth of mutant cells which are resistant to the effects of the dye on metabolism. The fact that pre-cancerous changes induced by carcinogens are non-reversible (see above) suggests strongly that selection of mutants is the correct explanation.

Most of the organic carcinogens are either capable of reacting with DNA (e.g. alkylating agents) or are metabolised within the cell to compounds, e.g. epoxides, which can do so. Such reactions are believed to occur between electrophilic groupings on the carcinogen and electron-rich groupings of DNA, notably in guanosine: but binding occurs at various sites within DNA, and may possibly result in the development of mutations of various types when the DNA replicates prior to mitosis. In view of the close correlation between carcinogenicity and mutagenicity (p. 299) *it seems very likely that chemicals induce cancer by causing mutations*. Such mutations can only occur in cells which undergo mitosis, and this may explain why cancer arises most commonly from labile populations of cells (p. 290), and also why mature neurons, which appear incapable of division, do not become neoplastic. In some instances where cancer arises in relatively stable cell populations, e.g. liver cells, cell loss and compensatory proliferation caused by disease, e.g. cirrhosis of the liver, or by the carcinogen itself, are probably essential for carcinogenesis.

It has also been suggested that chemical carcinogens may act by activating a latent oncogenic (tumour-producing) virus already present in the target cell (pp. 302 *et seq.*).

Physical agents in carcinogenesis

The main physical factor concerned with tumour formation is radiant energy, and much is known of this important form of carcinogenesis. Other physical factors, such as mechanical trauma, chronic irritation and the tendency for cancer to develop in scars, are difficult to study and their role in carcinogenesis is poorly understood.

Radiant energy. A detailed account of the effects of radiation on cells is included in Chapter 2. *Radiation of diverse kinds, x-rays, α , β and γ rays and ultraviolet light all induce tumours in man and animals.* They produce effects by release of energy during their passage through the tissues with resulting alteration of various cellular molecules, including the nucleic acids. The most important long-term consequence of these events is an increased rate of mutation in the irradiated cells. The degree of effect on a tissue depends largely on the total radiation dose, physical characteristics of the radiations (such as their penetrating capacity), and on features of the affected tissues, such as density, mitotic rate, and the nature of their blood supply. Thus short exposure to a high concentration of radiation, as occurred in those exposed to the atomic bomb or to accidents involving nuclear apparatus, and oft-repeated exposure to low doses of radiation, as occurred in the early radiologists, can both be carcinogenic. Tumours may arise in tissues affected by radiation necrosis (p. 38) and also in those where direct radiation injury appears to have been relatively slight.

The early **radiologists** frequently calibrated their machines by exposing their own arms. This cumulative exposure to x-rays was carcinogenic, but as the early x-rays were 'soft' and did not penetrate deeply through the skin, their tumours were accessible **squamous carcinomas of the epidermis** which could be successfully treated surgically. Later, however, despite the recognition of the hazard and introduction of safe practices, the use of 'harder', more penetrating x-rays led to a raised incidence of deeper tumours, and particularly to the development of **chronic myeloid leukaemia** originating in the haemopoietic marrow as a result of the capacity of bone matrix to impede and scatter the x-rays. Further developments in

safety techniques have now virtually abolished this problem.

The effects of **radio-isotopes** depend on the dosage, and the site of absorption of the radiation produced. Inhaled radioactive dust is likely to produce lung cancers (p. 497). Radio-iodine produces thyroid tumours in experimental animals because it is concentrated within the gland: external irradiation of the neck with x-rays can produce much the same effect.

The situation with **UV light**(UV) is interesting. *Long continued exposure to UV, especially medium-wavelength UVB, such as occurs in outdoor workers, is associated with the occurrence of basal cell and squamous carcinomas, which usually arise in those areas of skin exposed to light, and less often malignant melanomas of the skin.* Those most at risk are pale skinned individuals who tan poorly, albinos, and especially those with the genetic predisposition of xeroderma pigmentosum (p. 296). Not only is there a very high frequency of skin cancers in white inhabitants of sunny Australia (p. 297) but they affect relatively young people as compared with most other cancers, with some predilection for those of higher socioeconomic groups.

Much of the UV radiation in sunlight is filtered out by ozone in the stratosphere and does not reach the earth. There is current concern that the ozone is being reduced by supersonic transport and especially by fluorocarbons used as propellants in spray-containers and as refrigerants. It has been postulated that this will lead to an increase in skin cancer in sunny places, but the evidence on which this fear is based is, at present, inconclusive.

Trauma and chronic irritation. Many patients ascribe the onset of a visible (usually skin) tumour to some specific incident of trauma. It is difficult to understand how a gross mechanical injury could cause neoplasia, and it seems likely that, in most cases, an already growing tumour becomes more liable to injury and that this draws the patient's attention to the lesion—*traumatic determinism*. It is possible, however, that mechanical trauma may be important in a few cancers.

Chronic irritation, which could also be regarded as repeated minor trauma, seems im-

portant in squamous carcinomas of the mouth associated with ill-fitting dentures and in those cancers which arise in association with chronic infected sinuses opening onto the skin. Also, malignant melanoma in African negroes occurs mainly on the soles of the feet of those who go barefoot: this may not, however, be due simply to trauma, for most such tumours arise at the margin of the pale plantar skin, whereas trauma is obviously not so restricted in distribution.

Scars. Tumours do not ordinarily arise in a burn scar or surgical wound scar on a rabbit's skin. But if one paints a carcinogen evenly over an area which includes such a scar, the tumours

appear first (and grow largest) in relation to the scar. 'Burn cancers' of man, or 'brand cancers' of animals, occur almost always in areas exposed to excessive sunlight or similar carcinogenic stimulus. Scars thus seem to act as 'cocarcinogens'. Although chronic peptic ulcers of the stomach and duodenum have closely similar features, only those in the stomach appear to predispose to cancer. It may be that the chronic ulceration predisposes to cancer because of the active proliferation of the surface epithelium at the ulcer margins (which is where cancer develops) and that, in addition, the gastric (but not the duodenal) mucosa is exposed to carcinogenic agents.

Induction of tumours by viruses

Both RNA and DNA viruses have been shown to be oncogenic, i.e. capable of causing tumours, in various species of vertebrates. In many instances, however, it has not been easy to establish that a particular animal tumour is caused by a virus. The only human tumours which have been proved to be caused by a virus are the common wart and some venereal warts, but many of the procedures used to establish the viral nature of tumours in experimental animals are not applicable to man, and it is not surprising that evidence is only now accumulating which suggests that some human cancers are virus-induced. The evidence is, however, particularly strong for an unusual form of lymphoid cancer, the Burkitt lymphoma, and for nasopharyngeal carcinoma and fairly strong for liver cell cancer (p. 305).

Viral integration and transformation of host cells

When a virus particle enters a host cell, two things may happen. Firstly, in so-called *permissive cells* virus can **replicate**, producing infective virus particles which are released and may enter other cells. Secondly, sequences of virus DNA may be inserted into the DNA of the host cell, a process termed **integration**. *In all instances of virus-induced tumours which have been sufficiently elucidated, the development of cancer has been found to be associated with integration of*

the virus, and this phenomenon must be studied in some detail.

Many of the important advances in viral carcinogenesis have been made by investigating the effects of viruses on cell cultures, and some oncogenic viruses have been shown to be capable of **transforming** normal cells in culture into cells which behave like cultures of cancer cells. For example, cultures of normal cells proliferate only until such time as they form a continuous monolayer on the surface of the culture vessel: they then stop proliferating, a phenomenon known as **contact inhibition**. By contrast, virus-transformed cells (and cancer cells in culture*) continue to proliferate and become heaped up on one another (Fig. 11.6). Virus-transformed cells also show other features of cancer cells; for example increased motility, morphological changes and increased glycolysis, and in many instances they have been shown to develop into tumours when injected into histocompatible animal hosts: there is thus very strong evidence that certain viruses can convert normal cells into cancer cells *in vitro*.

The demonstration of integrated viral DNA in host cells has been achieved by elegant techniques, one of the most useful being *nucleic acid hybridisation*. To give an example, a suspected DNA virus is grown in a culture of permissive cells, which allow the virus to replicate,

*In general, cells from tumours of experimental animals grow much more readily than human cancer cells in culture, although in some instances the latter will grow continuously.

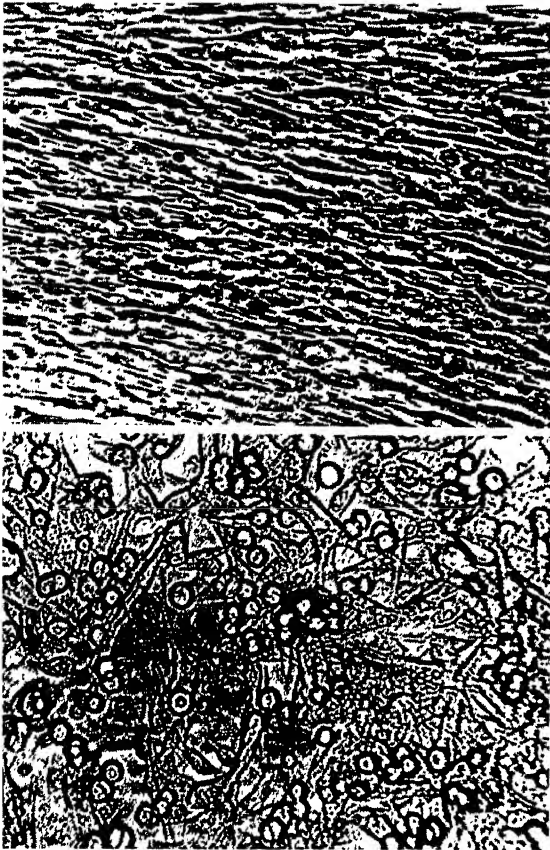
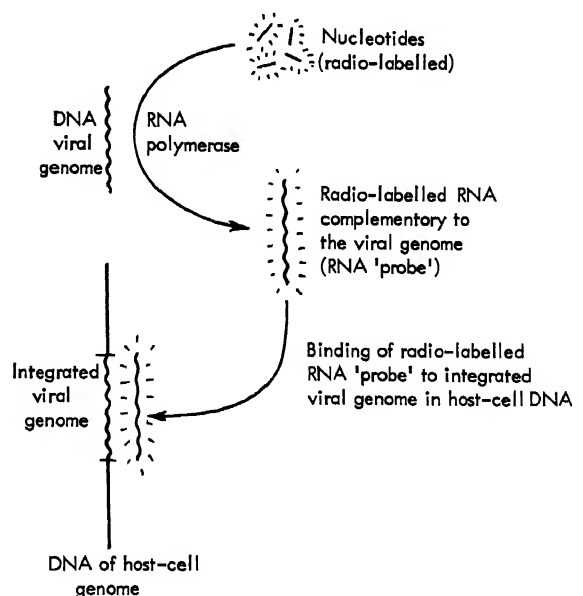


Fig. 11.6 Cultures of hamster fibroblasts (*left*). The upper culture shows formation of a regular monolayer. The lower culture is infected with polyoma virus and shows cellular pleomorphism and loss of contact inhibition, the cells being piled on top of one another. The upper hamster (*right*) was injected with the polyoma-transformed cells which have grown to form a tumour. The lower hamster was injected with the uninfected cells, and has remained healthy. (Dr. Joan McNab.)

and viral DNA is prepared in single-stranded form. Complementary strands of radioactive RNA are prepared *in vitro* by incubating the viral DNA with a mixture of suitable bases (one or more of which is labelled with a radio-isotope) and the enzyme RNA polymerase (Fig. 11.7). The labelled RNA thus produced will bind to the corresponding sequences of the viral DNA, and can be used to detect viral DNA integrated into the host cell genome. By this technique, viral DNA can be detected in various animal tumours and virus-transformed cells. The features of virus integration and its relation to tumour formation differ for DNA and RNA viruses, which are considered separately below.

Fig. 11.7 Use of the nucleic-acid hybridisation technique to detect integrated viral DNA. Note that the preparation of the RNA 'probe' requires replicating virus to provide the viral DNA genome, which is used as a template.



Oncogenic DNA viruses

In 1933, it was demonstrated by Shope that naturally-occurring benign epithelial tumours (papillomas or warts) of American cotton-tail rabbits could be transmitted by filtered extracts of the tumours. When scratched into the skin of cotton-tailed rabbits, benign tumours were produced. In domestic rabbits, the filtrates caused skin tumours which frequently became malignant. It is now known that oncogenic DNA viruses can *either* integrate and transform a cell, *or* replicate within it: they cannot do both in the same cell, for replication of virus kills the cell. The infectivity of the cotton-tail papillomas was due to replication of the virus (subsequently termed the *Shope papilloma virus*) in keratinised, non-dividing epithelial cells: in the proliferating tumour cells, the virus is integrated, but does not replicate. The cells of the more malignant tumours produced in domestic rabbits contain integrated, but not replicating virus, and their filtrates are non-infective. This point is important, because in other DNA virus tumours, infective virus is not produced, and integrated virus can only be detected by nucleic-acid hybridisation (see above) or by means of **cell hybridisation**. The latter consists of fusing together two cells from animals of the same or different species, forming a tetraploid cell. When a (non-permissive) cell containing an integrated DNA virus is fused with a permissive cell, complete virus particles may be formed in the hybrid cell, thus revealing the presence of the integrated virus in the non-permissive cell. Skin papillomas in other species have been shown to be caused by papilloma viruses, including the human common wart (*verruca vulgaris*) and the venereal wart (*condyloma acuminatum*); in both these human tumours, the virus replicates in some non-dividing cells, and so they are infectious.

The papilloma viruses belong to a group termed *papovaviruses*, which include also *Polyoma virus* and a virus termed *SV40*. *Polyoma virus* is unusual in that it can induce tumours of various types (hence the name) in several species: most oncogenic viruses are more specific in the type of tumour and in the species in which they induce tumours. Polyoma virus occurs naturally in mice, and when injected into neonatal mice, other rodents or rabbits, it induces a wide range of tumours. When added

to cell cultures, it integrates into and transforms some cells and replicates in others. The relative ease with which Polyoma virus produces tumours when injected into *neonatal* animals is a general feature of oncogenic viruses: when introduced into adult animals, the host immune response usually results in antibodies which inactivate free virus, and in cell-mediated immunity which destroys any transformed cells. For this reason, tumours are rare in the naturally-infected adult mouse, but tumours can be produced by introducing the virus into immunologically suppressed or congenitally immunodeficient (nude) mice. *SV40* occurs naturally in monkeys, in which it replicates but does not produce tumours, but it can cause tumours when injected into neonatal rats and can transform cells in culture. This illustrates a general feature of DNA oncogenic viruses—they usually replicate and destroy cells in the natural host and integrate into cells of the host(s) in which they are oncogenic. Perhaps because of man's close relationship to monkeys, *SV40* can infect man, producing in immunosuppressed or immunodeficient patients a form of encephalitis by replicating in cells of the brain; it has not, however, been shown to produce human tumours.

Herpesviruses

Two herpes viruses are of considerable interest because there is evidence associating them with tumours in man. One, *Epstein-Barr virus*, is associated with a tumour termed the Burkitt lymphoma and also with cancer of the nasopharynx: the other, *Herpes simplex virus type 2*, has been associated with carcinoma of the uterine cervix.

Epstein-Barr virus (EBV). In certain parts of Africa where malaria is highly endemic, Burkitt noted the relatively high incidence of an otherwise rare tumour of lymphocytes which occurs in children and affects the jaw and various internal organs. Cells from this tumour, the **Burkitt lymphoma**, are lymphoblasts derived from B lymphocytes and grow continuously in cell culture. Occasional cells in such cultures produce EBV particles and this was the original source of the virus. Most of the cells in culture, and those of the tumour itself, have been shown to contain several EBV genomes in their nuclear DNA. Moreover, when added to cul-

tures of human lymphocytes, EBV transforms them into lymphoblasts resembling those derived from the tumour. Involvement of B lymphocytes appears to depend on their possessing a surface receptor for C3d (an activation product of the third component of complement) which also binds EBV and so facilitates its entry into B lymphocytes.

EBV is widespread throughout the world. Most people develop immunity to it without clinical illness, but in some it causes the acute febrile illness known as *infectious mononucleosis*: in this, antibody to EBV appears and there is also an unusually vigorous immune response on the part of T cells, which presumably eliminate any B cells transformed by the virus. Complete recovery is the rule and there is no evidence of subsequent increased risk of developing a lymphoma. The occurrence of the Burkitt lymphoma in Africa is associated closely with endemic malaria, and the tumour incidence is low in areas where malaria has been eradicated and in Africans with the sickle-cell trait of their red cells, which protects against malaria. The most likely explanation of the association with malaria is the effect this disease has in depressing cell-mediated immunity, with consequent failure to destroy lymphocytes transformed by EBV. The suppressive role of immunity is supported by the spontaneous regression of the tumour in some cases, and by its unusually good response to low doses of anti-tumour drugs.

EB virus is also associated with poorly differentiated carcinoma of the nasopharynx, which is the commonest form of human cancer in parts of Southern China. Integrated EB virus DNA can be detected in the tumour cells. The high incidence of this otherwise rare tumour in the Southern Chinese is probably due to a genetic factor, for it occurs also in Chinese groups who have settled in other countries, and is rarely observed in Caucasians living in Southern China. Recently, a high incidence of the histocompatibility antigen HLA-A2 has been reported in patients with this cancer, and it may be that this is associated with a particular repertoire of immune response (Ir) genes (p. 167) which do not provide for a normal vigorous T-cell response to EBV-transformed epithelial cancer cells.

Herpes simplex type 2 virus (HSV2) is known to be transmitted sexually, and female genital

infection is associated with sexual activity and promiscuity. These behaviour patterns are also associated with a high incidence of carcinoma of the cervix uteri, and patients with this condition have higher titres of antibody to *HSV2* than those without the disease. *HSV2* has also been detected in cells from cervical cancer. Although these findings are suggestive, it still remains entirely possible that *HSV2* infection is incidental and that the virus is simply a 'passenger' and not the cause of the tumour.

Other DNA viruses

Hepatitis B virus is associated with acute and chronic hepatitis in man. Chronic hepatitis commonly progresses to cirrhosis of the liver, which in a proportion of cases is followed by cancer of the liver cells. A factor of importance in this progression is the continuous loss and compensatory proliferation of liver cells, which are normally stable (p. 77). As mentioned earlier, such proliferative activity predisposes to cancer. There is, however, evidence that hepatitis B virus may be more directly carcinogenic, for liver cell cancer is a much less common complication of cirrhosis due to causes other than the virus, e.g. chronic alcoholism. Also, there is preliminary evidence suggesting that hepatitis B viral DNA is integrated into the genome of liver cancer cells. Experimental animal studies have proved difficult with this virus, and more evidence is needed to determine its role in human cancer. The subject is discussed more fully on pp. 701–2.

Adenoviruses are a common cause of upper respiratory infections in man. One of them (type 12) has been shown to be capable also of causing tumours in rodents and of transforming rodent cells in culture. There is, however, no evidence that they are oncogenic for man.

How do oncogenic DNA viruses cause cancer?

Integration of viral DNA into a host cell could produce the transformation to a cancer cell either by its direct effect on the host-cell DNA, i.e. by acting as or causing mutations. Alternatively, the virus might contain a gene coding for a product which transforms the host cell. Some support for the mutation theory has been provided by the demonstration that although a transformed cell may contain several genomes

of the transforming virus in its DNA, transformation is associated with integration into a particular chromosome. This does not, of course, prove that a particular mutation occurs and is responsible for transformation, and indeed there is evidence that transcription produces a protein which is necessary for both integration and transformation of the host cell. This protein, termed *T-antigen*, is detectable in the nucleus of transformed cells, and may induce transformation by binding to host DNA. Antibody to the appropriate T-antigen, which is specific for each oncogenic DNA virus, is found in the serum of animals with DNA-virus induced tumours. Some of the oncogenic DNA viruses have only sufficient DNA to code for a few proteins, and elegant studies have been performed in which the small DNA strands of such viruses have been broken into several fragments by means of enzymes. The individual parts of the DNA may then be isolated and their effects tested on suitable host cells in culture. In such experiments, it has been shown that *only the part of the viral DNA which codes for T antigen is capable of integrating into and transforming host cells.*

Oncogenic RNA viruses (retraviruses)

Introduction. RNA viruses which produce tumours are commonly termed *oncornaviruses* (an abbreviation of oncogenic RNA viruses) but increasingly they are described as *retraviruses* because of their capacity to produce reverse transcriptase (see below).

Like the DNA oncogenic viruses, the retraviruses must become integrated into the host cell to induce transformation to a cancer cell. To do so, the retravirus, having invaded the host cell, synthesises a strand of DNA complementary to its RNA genome, by means of *reverse transcriptase* (RNA-directed DNA polymerase). The viral-coded DNA, known as a *provirus*, is then inserted into the host cell genome (Fig. 11.8) Such integration is restricted mainly to RNA viruses which have been shown to be oncogenic, and is essential for viral transformation to a cancer cell. The integrated DNA provirus can be transcribed into the RNA genome of the virus, and (unlike other RNA viruses) replication occurs only in this way. Nearly all retraviruses have a char-

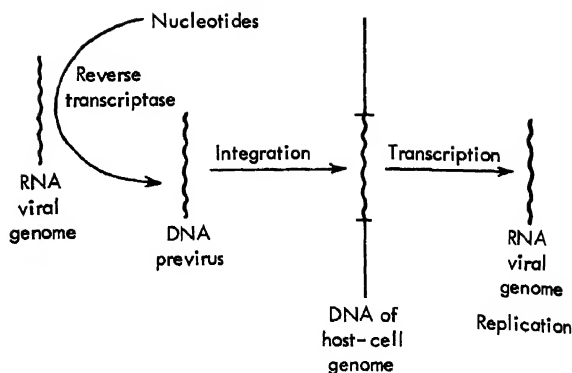


Fig. 11.8 Formation of an RNA viral provirus and integration into the host-cell genome. Viral replication is effected by transcription of the integrated provirus.

acteristic morphology: they form *C-type particles* at the host cell membrane by a process of 'budding,' and the infective virion thus includes part of the host cell membrane (Fig. 11.9). This form of replication can occur without killing the host cell and thus, unlike DNA viruses, the retraviruses can both integrate and replicate in the same cell.

The presence of an integrated oncogenic RNA virus in cells may be detected by means of nucleic acid hybridisation (p. 302), using a transcript of the viral RNA genome as a template for production of a DNA 'probe'. Strong evidence of the presence of integrated virus is provided also by the detection of reverse transcriptase, and also by the appearance of provirus-coded proteins on the surface of the host cell (p. 314).

Types of oncogenic RNA viruses

There are three main groups of oncogenic retraviruses—the *sarcoma viruses*, *leukaemia viruses* and *mammary tumour viruses*.

RNA sarcoma viruses

The first of these was discovered by Rous in 1911, when he showed that a sarcoma of chickens was transmissible by a filtered extract of the tumour. This agent (*Rous sarcoma virus*) and other similar avian-tumour producing viruses rapidly transform chicken fibroblasts in culture. Some strains of sarcoma viruses are *deficient* in that they can transform cells but lack some of the genetic material required for replication: they may be induced to replicate by



Fig. 11.9 A part of a cell from a cat infected with feline leukaemia virus, showing formation of a virion by budding from the cell surface. Note that the envelope is formed from the plasma membrane, and that the virus particle becomes coated with an outer spiky layer. The section happens to include part of another cell immediately above the virion. $\times 80\,000$. (Dr. Helen Laird.)

addition of a 'helper' virus which supplies the defective gene product, and some defective viruses will replicate in cells which already contain an *endogenous virus* (i.e. an integrated virus transmitted genetically from generation to generation by the germ cells—p. 309) which makes good the deficiency. Sarcoma viruses affecting a number of mammalian species have been described.

RNA leukaemia viruses

Viruses of this group induce various types of cancer of lymphoid cells (lymphomas), including leukaemia, in chickens, mice, cats and bovines. For convenience, this group of tumours will be referred to as *leukaemia* in this

account. Chicken leukaemia virus was discovered as long ago as 1908, when chicken leukaemia was shown to be transmissible, most readily to newly-hatched chickens, by a filtrate of leukaemic cells. Mouse leukaemia virus was demonstrated in the leukaemic cells obtained from inbred strains of mice in which 'natural' leukaemia was common. Filtrates of the leukaemic cell extracts were shown to induce leukaemia when injected into neonatal mice of strains in which leukaemia was uncommon. It is noteworthy that leukaemia virus in replicating form is detectable in the cells of *all* fetal mice, and it has even been speculated that it plays a physiological role in the proliferation and differentiation of fetal cells. During maturation, replicating leukaemia virus disappears, but integrated virus persists in the cells of various tissues, and may be detected by nucleic-acid hybridisation, or by treating mice with x-rays or various chemical carcinogens, which results in replication of the virus, and also in leukaemia. In high-incidence strains the spontaneous development of leukaemia is also associated with the reappearance of replicating virus, probably because the latter impairs the T-cell immune response of the animal and so allows transformed lymphocytes to proliferate.

Feline leukaemia virus, the cause of leukaemia in cats, is of considerable interest for at least four reasons. Firstly, infection by feline leukaemia virus occurs naturally by horizontal transmission (i.e. by personal contact); secondly, the virus spreads among an *outbred* cat population; thirdly, infection of adult cats can induce leukaemia, and fourthly it has proved possible to protect cats from leukaemia by use of a vaccine which is effective even when administered *after* the animal has become infected by the virus. These findings are due largely to the work of Jarrett and his co-workers in Glasgow. They have shown that, while vertical transmission of the virus does occur, as in mice, horizontal transmission is usually responsible for leukaemia. Young animals receiving a large infecting dose of the virus are more likely to develop leukaemia, probably because these factors lead to immunosuppression, and epidemiological studies have shown that cats of a low 'social class', which roam freely, are much more likely to become infected and to develop leukaemia.

Viruses and human leukaemia. It is apparent

that there are important differences between virus-induced leukaemia in different species, and this must be taken into account in elucidating the causal factors of lymphomas and leukaemias in man. Retroviruses, including the leukaemia viruses, can replicate in cells in culture, but do not transform them. In the living animal, integrated leukaemia virus may remain latent (i.e. may give rise to no detectable products) throughout life although, as noted above, it may be activated by administration of carcinogens or x-rays. In man, such latent proviruses, if present, would be extremely difficult to detect, although the development of leukaemias in people exposed to heavy doses of ionising radiations could be explained by activation of an endogenous leukaemia virus. Antibodies to leukaemia viruses are very commonly present in human serum, and C-type virus particles have been observed in some cultures of human leukaemic cells, while nucleic-acid hybridisation techniques have revealed the presence of apparent leukaemia proviruses in such cells. There is thus some preliminary evidence that, as in a number of animal species, certain forms of leukaemia in man may be due to retroviruses. Indeed, in view of the demonstration that retroviruses cause leukaemia in several species, and that feline leukaemia virus can induce a variety of lymphomas closely resembling most of those observed in man, it would be surprising if leukaemia viruses were not the cause of some types of human lymphomas.

A major difficulty in the detection of oncogenic viruses in human lymphomas and leukaemias is the lack of a specific 'probe' for the provirus. The base sequences in feline and bovine leukaemia viruses show considerable differences, so that bovine leukaemia provirus cannot readily be detected by nucleic acid hybridisation with a probe prepared from feline leukaemia virus RNA. Accordingly, while leukaemia viruses are likely to be involved in human tumours, the detection of provirus requires the preparation of RNA from replicating virus, and this has not yet been accomplished. It must also be emphasised that, in feline and bovine leukaemia, the causal virus can be detected in only about 50% of tumours, although from epidemiological studies it is very likely that the 'virus-negative' tumours are caused by leukaemia viruses.

Mouse mammary tumour virus is responsible

for breast cancer in mice. It is of interest for three reasons. Firstly, it accounts for a cancer which originally appeared to be determined genetically. Secondly, the virus causes tumours particularly in female mice, and requires oestrogen stimulation of the target epithelial cells. Thirdly, there is some evidence that a similar virus may be involved in human breast cancer. By selective breeding, strains of mice were established in which virtually all the females developed breast cancer late in life, and other strains in which this tumour was rare. This looked like a genetic effect until it was found by Bittner that female neonates of a low-cancer strain suckled by foster mothers of a high-cancer strain frequently developed breast cancer. By contrast, high-cancer strain offspring suckled on low-cancer foster mothers did not develop cancer (Fig. 11.10). Thus the carcinogenic influence was transmitted by the milk in the post-natal period and not by germ

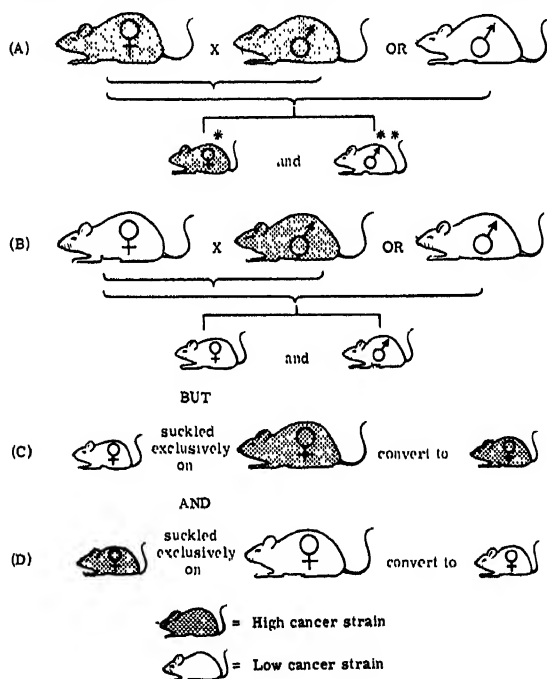


Fig. 11.10 Discovery of the Bittner milk factor. The incidence of breast cancer was high in the daughters of 'high-cancer' strain mothers (A), but low in the daughters of 'high-cancer' strain fathers (B). The strain of female on which neonates were suckled (C and D) was then found to determine the incidence of cancer, showing the importance of a milk factor, since shown to be a virus. Oophorectomy* reduces the incidence of breast cancer in virus-infected females, and oestrogen** increases the incidence in virus-infected males.

cells nor transplacentally by the mother. Investigations with filtrates of 'high-cancer strain milk', cell-free filtrates of mammary tumour, and electron microscopic studies of milk and tumour tissue, have confirmed that the active principle ('Bittner milk factor') is a virus—mouse mammary tumour virus.

Female neonates which have been infected with the virus are protected from breast cancer by removal of the ovaries, and male neonates infected with the virus only develop breast cancer if given large doses of oestrogens. Thus oestrogenic activity is a necessary co-factor for this virus-induced tumour.

How do reoviruses cause cancer?

Integration of provirus into the host cell genome is an essential feature of cancer induced by reoviruses, but this alone is not sufficient to transform the cell to a cancer cell, for it is known that in many vertebrate species integrated reoviruses are transmitted genetically via the germ cells (*endogenous viruses*) and are thus present in all nucleated cells. In some instances, it is probable that expression (i.e.

transcription) of one or more genes of the provirus, is necessary to transform the host cell. The mechanism of such activation of viral 'oncogenes' is not known, although, as noted earlier, administration of chemical carcinogens and ionising radiation may induce changes which trigger it off. In the case of the sarcoma viruses, which transform cells in culture, transformation has been shown to be effected by a single oncogene. This may be detected by nucleic-acid hybridisation, and its continued expression appears to be necessary for both transformation and maintenance of the transformed state in the host cell. Its product is a protease capable of acting on several cellular substrates, including components of the cell membrane.

Information on the mechanism of induction of cancer by leukaemia viruses is scanty. There is, however, preliminary evidence that it may depend on a proviral gene which behaves as an operator, inducing transcription of an adjacent gene of host (i.e. non-viral) origin. If this is correct, then it appears that the sarcoma and leukaemia viruses induce cancer by different mechanisms.

Hormones and carcinogenesis

In general, any induced change of hormone level which causes prolonged hyperplasia of a target organ may cause tumours of the latter, but there are many exceptions. The following are the best known experimental situations.

Oestrogens can undoubtedly cause tumours in susceptible strains of mice; their administration in high dosage leads to an increased incidence of cancer of the breast in females and to the occurrence of breast cancer in males. Reduction of natural oestrogen levels by oophorectomy abolishes cancer of the breast in susceptible female mice. It might seem that the excessive proliferation of breast ducts induced by oestrogen, carried to excess, has been the actual cause of the cancer. But, as noted above, the oestrogens appear to act effectively only in the presence of the mammary tumour virus in mice. In virus-free mice (and in other species) the effect is much harder to demonstrate. In tissues other than the breast the position

becomes somewhat anomalous. The most obvious oestrogen target cell, the endometrial glandular epithelium, rarely develops tumours in treated animals, though connective-tissue tumours of the uterus are often produced and tumours result also in organs not usually regarded as oestrogen-responsive, for example the kidney in the hamster, and the Leydig cells of the testis in the mouse. These effects appear to depend only on the oestrogen activity of the various compounds concerned, and not on their precise structure.

It now seems likely that there is an increased risk of carcinoma of the endometrium in women receiving prolonged oestrogen therapy and in patients with an oestrogen-secreting granulosa-cell tumour of the ovary. In the past the risk has been exaggerated by confusion between endometrial hyperplasia and carcinoma.

Contraceptive hormonal preparations. Considering the very large number of women

taking oral contraceptive pills there is little evidence of any carcinogenic effect. There is a small increase in benign tumours of the liver, which correlates with dose and duration of therapy, while benign breast lesions are decreased. There is no obvious change in the incidence of cancer of the breast or uterus. An early type of 'pill', in which different hormones were administered sequentially, was associated with an increase in endometrial carcinoma, and has now been withdrawn.

Androgenic/anabolic steroids. These hormones, notoriously used by athletes competing in field events to increase muscle mass, may be involved in the development of cancer of the liver.

Experimental endocrine disturbances and tumours. Experimental procedures which induce an increased output of trophic hormones by the adenohypophysis have been shown to result in cancer in the target organs, although trophic hormones have not been shown to induce cancer in man.

Examples of this mechanism of tumour induction include the following. (a) If the ovaries of a rat are removed and pieces are implanted into the spleen, they continue to secrete oestrogen, but this passes via the portal vein to the liver, where it is mostly inactivated. In consequence, there is increased secretion of FSH by the adenohypophysis and a granulosa-cell cancer eventually develops in the stimulated follicular tissue of the transplanted ovaries. (b) If rats are treated with a drug such as thiouracil, which blocks the production of thyroid hormone, increased secretion of TSH by the adenohypophysis causes hyperplasia and eventually cancer of the thyroid follicular epithelium. Cancer develops more rapidly, and with more certainty, if a carcinogen, e.g. 2-acetylaminofluorene or radio-iodine (which is taken up by the thyroid epithelium) is administered to the experimental animals.

Another example of functional hyperplasia leading to neoplasia is provided by removing the thyroid gland in mice, or destroying it with a large dose of radio-iodine. In the absence of

thyroid hormone, the TSH-secreting cells of the adenohypophysis undergo hyperplasia and in some strains of mice this progresses to cancer. It is of interest in relation to the following section that initially the cancer cells can be suppressed by thyroxine, but eventually they may continue to grow when transplanted serially into mice with normal thyroid function.

Hormone-dependent tumours in man. The pituitary and thyroid tumours just mentioned may both be 'hormone-dependent' in the sense that they may regress if the hormonal disturbance that invoked them is corrected. Related phenomena in man are few, but the following three carcinomas deserve mention. (1) Many **prostatic carcinomas** are sufficiently dependent on a normal male hormonal environment to be slowed down, arrested, or even to regress for long periods if oestrogens are given. (2) Some differentiated **thyroid carcinomas** are partially dependent on TSH, and their rate of growth and spread may be reduced or arrested by continued administration of thyroxine, which suppresses secretion of TSH by the pituitary. (3) Some **breast carcinomas** regress under various hormonal manipulations—treatment with male hormones or even oestrogens, oophorectomy, adrenalectomy, hypophysectomy. Treatment by these methods has been largely empirical, but the cells of some breast cancers have receptors for oestrogens or other steroid hormones, such as progesterone. When hormone binds to the receptors the hormone-receptor complex enters the cell and is passed to the nucleus where it affects nucleic acid metabolism. The clinical significance is that patients with tumours consisting of receptor-positive cells are likely to respond to hormone therapy, whereas those whose tumours are receptor-negative are unlikely to respond.

Just as the experimentally-induced hormone-dependent pituitary tumours become independent after serial transplantation (see above), hormone-sensitive tumours in man practically always ultimately resume growth, though with the thyroid and prostatic carcinomas the period of arrest or partial regression is often long.

Epidemiological considerations in cancer prevention

There are two major approaches to prevention of cancer. One is by basic biological research on the mechanisms of the cellular changes leading to the development of cancer in animals and man. The second is by detecting carcinogenic factors in the environment, i.e. in the air, soil, foods and in industrial processes, and taking steps to eliminate exposure to them.

So far, we have considered the roles of genetic factors, chemical and physical agents, viruses and hormones in the development of cancer. We must now discuss the epidemiological approach to the detection of environmental factors which, together with the genetic predisposition or resistance to particular forms of cancer, determine the incidence and types of cancer to which man is subject. This is an exceedingly complex subject, and the following account is necessarily limited to some general principles and illustrative examples.

Life expectancy. As illustrated in Fig. 11.11, the risk of cancer increases with age. In this country, the life expectancy of newborn infants has increased from 40 to 70 years in the last 100 years, and this has been accompanied by a very great increase in cancer, which is now the cause of death of just over 20% of the popula-

tion. By contrast, in countries with a much shorter life expectancy (due mainly to malnutrition and infections) the incidence of cancer is much lower. One reason for this relationship between cancer and age is obvious: *the cellular changes leading to cancer progress slowly over a long period, and accordingly cancer usually develops many years after first exposure to environmental carcinogens.*

Natural environmental factors. The importance of U.V. in *sunlight* as a cause of skin cancers, particularly in pale-skinned people, has already been considered (p. 297). Other natural factors which may have an influence on the incidence of cancer include *geological features* which influence the level of *background radiation* and also the amounts of various chemicals in the soil, water and vegetable foods. Climatic and other factors also determine the presence and incidence of certain parasitic diseases which predispose to cancer, e.g. schistosomiasis (bladder cancer in Egypt) and liver flukes (biliary-duct cancer in the Far East).

Occupational factors. Some examples of cancer resulting from exposure to carcinogens used in industry have already been given (pp. 298–301); these and some others are listed in Table 11.3, but it should be emphasised that, with the ever-increasing variety of industrial processes, the number of chemical carcinogens and co-carcinogens is now enormous, and screening of compounds potentially useful in industrial processes or as drugs, food additives, cosmetics, weedkillers, insecticides, etc., is becoming an increasingly heavy task. As already pointed out (p. 299) it may take years to detect that a substance is carcinogenic in animals, and even the most careful tests on animals cannot exclude the possibility that a substance is carcinogenic to man. Nor does it always follow that a substance found to be carcinogenic in the laboratory will be a danger to man. It is therefore important that general medical practitioners and epidemiologists should be on the alert for an unusually high incidence of cancer in particular occupational groups. A good example is provided by bladder cancer in workers in the aniline-dye and rubber industries (p. 299). More recently, it has become clear that asbestos miners and workers inhaling asbestos

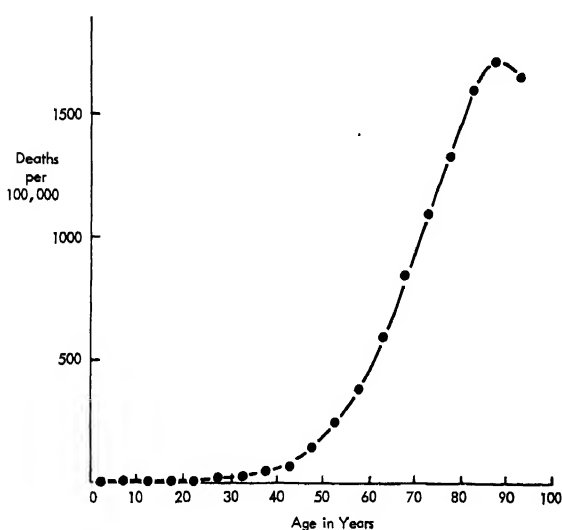


Fig. 11.11 Deaths from cancer in England and Wales during 1977, shown as the number of deaths per 100 000 for each 5-year age group. (Mortality Statistics, 1977, HMSO, London).

Table 11.3 Some Examples of Cancer Associated with Occupational Exposure to Carcinogens

Types of tumour	Occupational group	Carcinogen
Skin cancer	(a) Radiologists (b) Farmers and fishermen	(a) 'Soft' x-rays (b) Ultraviolet light
Chronic myeloid leukaemia	Radiologists	'Hard' x-rays
Bladder cancer	Rubber workers Aniline dye workers	1-hydroxy- 2-naphthylamine
Lung cancer	(a) Miners (b) Uranium miners	(a) ? Silica (b) γ -radiation
Mesothelioma of pleura or peritoneum	Asbestos miners Pipe-laggers, etc.	Asbestos
Adenocarcinoma of nasopharynx	Furniture makers	Wood-dust
Angiosarcoma of liver	Polymerisation-chamber cleaners	Vinyl chloride monomer

dust, for example in the manufacture of vehicle brake linings and heat insulating material, have an unusually high risk of developing pleural or peritoneal mesotheliomas and bronchial carcinoma. In the plastics industry, an unusually high incidence of an otherwise rare tumour, haemangiosarcoma of the liver, has been observed in workers producing polyvinyl-chloride (PVC). This plastic is not itself carcinogenic, but it is manufactured by polymerisation of vinyl chloride monomer (VCM) which is carcinogenic. The risk is apparently greatest for workers who clean out the polymerisation tanks in which relatively high concentrations of residual VCM are inhaled, and safety measures have accordingly been introduced.

The risk of industrial carcinogens may spread beyond those directly involved in their use. For example, the spouses of asbestos workers have been reported to have an increased risk of asbestos-related cancers, and people living close to industrial plants emitting solvent vapours (which may contain ketones, alcohols, aromatic hydrocarbons, ethers, chlorinated hydrocarbons and nitrites), smoke or dust, are believed to have an increased incidence of lymphoid and other cancers. A recent study of the distribution of respiratory tract cancer in a Scottish town revealed a higher incidence among those living downwind from a steel foundry. Analysis of the polluted atmosphere showed relatively high concentrations of sulphur dioxide and of dusts containing compounds of iron, manganese, nickel, lead and cadmium. Similarly, quite apart from the hazard of cigarette smoking, the incidence of

lung cancer is significantly greater among those living in industrial zones than in rural communities.

Social factors. The countless ingredients of food, use of an ever-increasing number of drugs and various social habits are without doubt of importance in the causation of cancer. Dietary factors are discussed below. Two social habits known to cause cancer are cigarette smoking and betel chewing. *Cigarette smoking* is very largely responsible for the high incidence of lung cancer throughout the world: it accounts for the death, from carcinoma of the bronchus, of approximately 10% of men over 45 years old in this country (i.e. approximately 40% of male cancer deaths). The chewing of *betel leaves* mixed with tobacco leaves and slaked lime is widespread in Southern India and South East Asia, and is associated with a high incidence of cancer of the oral mucosa. It is a depressing fact that, although the consequences of these two habits have now been known for many years, they continue to be largely disregarded.

The influence of social habits is illustrated also by the incidence of *breast cancer*, which is lower in women who have borne children than in nullipara and appears to be lower in mothers who have suckled their children than in those who have used artificial infant foods (but see p. 979). A curious observation was the higher incidence of cancer in the left breast than in the right among women of the Tanka boat people in Southern China. This is possibly due to the design of the tunics they wear, which makes it easier to feed their infants from the right breast. In experimental animals, breast duct ligation or excision of the nipple is followed by

an increase in cancer in the breast so treated, and it is postulated that retained breast secretions have a carcinogenic effect.

Dietary factors. The wide variations in diet, including the nature and amounts of foods eaten, methods of cooking, and the types of cooking vessels used, are all likely to influence the incidence of cancer, particularly of the alimentary tract. Additional potential sources of carcinogens in food include chemical fertilisers, insecticides, organic and inorganic chemicals fed to livestock, and chemicals used in the preservation of food. Natural diseases of plants and animals used as food may also be of importance, as exemplified by the experimental production of liver cancer by feeding aflatoxin, a product of the fungus *Aspergillus flavus* which contaminates ground nuts.

Investigations on *carcinoma of the large intestine* (colorectal cancer) illustrate how epidemiological and laboratory studies may help to elucidate causal factors in cancer (Hill, 1977). The incidence of this tumour has been shown to be relatively high in technologically advanced countries with a high standard of living. The incidence in migrant groups who have adopted the dietary habits of their new country changes from that of their country of origin to that of their country of adoption and thus genetic factors appear to be unimportant. The incidence is approximately the same in both sexes, and hormonal factors are unlikely to be involved. As regards diet, the incidence shows a positive correlation with the total caloric intake and with the amount of meat, animal fat and protein in the diet, and a negative correlation with the amount of vegetable fibre: animal foodstuffs are thus in some way responsible. Although many dietary ingredients have been suspected, current interest is centred on the production of carcinogens or co-carcinogens by bacteria in the lumen of the large intestine. It has been shown that gut bacteria can convert dietary factors to substances carcinogenic for animals, for example methylazoxymethanol from β -glucoside and N-nitrosamines from secondary amines and nitrate. Thus a diet which provides a suitable substrate and en-

courages the colonisation of the colon by particular species of bacteria might lead to the production of carcinogenic bacterial metabolites in the gut. A number of such possibilities exist: for example, gut bacteria can produce carcinogenic or co-carcinogenic metabolites from a variety of compounds present in food or produced by digestion in the gut, including tryptophan, tyrosine, methionine, cycasin and cholesterol. The available epidemiological and other evidence does not, however, provide strong support for an important role of these various bacterial metabolites in colorectal cancer. Nor can the geographical distribution of this tumour be explained readily by the carcinogenic activities of N-nitrosamines, aflatoxin (see above) or polycyclic aromatic hydrocarbons in the diet (although the latter may well play a causal role in gastric cancer).

At present, there is considerable interest in the possibility that *bacterial metabolites of bile acids* may be of importance. A correlation has been shown between the mean concentration of total bile acids in the faeces and the incidence of colorectal cancer in the population, and analysis of the faeces of patients with this tumour has shown a higher mean concentration than that of controls. A group of bacteria known as the nuclear dehydrogenating clostridia (NDC) are capable of desaturating the bile acid nucleus, producing unsaturated bile acids which have been shown in animal studies to act as carcinogens or co-carcinogens in the large bowel. These bacteria are commonly present in the faeces of populations with a high incidence of colorectal cancer and rare where the incidence is low: they are particularly common in the faeces of patients with colorectal cancer. The effect of diet may thus be to favour a flora which includes bacteria capable of converting bile acids to carcinogens. It is noteworthy that in communities with a high incidence of colorectal cancer most of the tumours develop in the left side of the colon and rectum, and the protective action of a diet rich in *vegetable fibre* may be attributable either to its effect on the bacterial flora or to more rapid transport of the gut contents through the lower part of the large intestine.

Host reactions in cancer

The natural history of cancer is not determined solely by the characteristics of the tumour cells, but also by the host's reaction to them. As indicated below, there is good evidence that protective host mechanisms exist, and that although these are very often unsuccessful in preventing the growth and spread of malignant tumours, and only very rarely bring about their complete destruction, they may nevertheless restrict the rate of tumour growth and spread and contribute to the degree of success achieved by various forms of treatment.

The nature of the host defences is largely unknown, but it has been shown that the cells of human and animal tumours possess surface antigens which are sufficiently foreign to the host to stimulate an immune response. This important property of tumour cells raises the possibility of immunotherapy, and in consequence there is considerable interest in the immunology of cancer, some of the major features of which are summarised below.

Experimental animal studies

Much of the experimental work has involved transplantation experiments, and before the importance of 'transplant' alloantigens was appreciated, rejection of transplanted tumours was frequently observed, but is likely to have been due to histo-incompatibility rather than to a tumour-specific reaction. Since the importance of alloantigens was demonstrated in mice by Gorer, the provision, by close inbreeding, of syngeneic strains of mice and rats has greatly facilitated experimental cancer research, not only by eliminating the 'transplant' antigens as a cause of rejection, but also by excluding other genetically-determined variables.

Specific immune responses to tumours

The elimination of cancer cells by a specific immune reaction on the part of the host requires (1) that cancer cells exhibit antigens to which the host is capable of mounting an immune response, and (2) that the products of the host immune response—antibodies and/or specifically reactive T lymphocytes—are capable of effecting destruction of the tumour cells. If these conditions are fulfilled, it must further

be asked why tumours grow and spread in spite of the host's immunity, and whether the balance can be tilted in favour of the host.

Tumour-cell antigens. To render a tumour cell susceptible to an immune reaction, the tumour cell antigens must be exposed on the cell surface, and must also be sufficiently foreign to stimulate an immune response; this excludes those surface antigens which, although present on tumour cells, occur also on the host's normal cells. Thus species- and organ-specific antigens, 'transplant' antigens (e.g. those of the H2 system in mice and of the HLA system in man) and blood-group antigens are all found (although often in reduced concentration) on tumour cells, but none of them is tumour-specific. There are, however, surface antigens on tumour cells which cannot be detected on normal cells: they are capable of eliciting immune responses and reactions and are sometimes called tumour-specific transplantation antigens (TSTA). In experimental carcinogenesis, these apparently tumour-specific surface antigens are largely dependent on the agent which has induced the tumour. *All tumours produced by any one oncogenic virus have been found to have relatively strong common surface antigens.* These include antigens which, although coded for by the viral genome, are not a structural component of the virion, and also, in the case of C-type viruses, antigens of the virus envelope which consists of modified host-cell-membrane (p. 306). It follows that antibody or primed T-lymphocytes reactive with these virus-coded antigens on a particular tumour should be reactive with the cells of all tumours produced by the same virus. By contrast, *tumours induced by chemical carcinogens or by physical agents have only weak common antigens; they develop stronger tumour-specific surface antigens, but these differ for each tumour, even when multiple tumours have been produced in the same tissue of the same animal by the same carcinogen.*

Immune responses to tumour-specific antigens. Both antibodies and cell-mediated immunity have been demonstrated, by *in-vitro* techniques, to develop in animals bearing tumours. In general, these responses are more readily demonstrable when the tumour is small, and as it enlarges and spreads they tend to diminish and

disappear. They are readily demonstrable after excision of the tumour and then gradually diminish unless the tumour recurs or is re-introduced into the animal.

Effects of the immune response to tumours. Although tumours grow in spite of the host immune response, under experimental conditions a protective effect can sometimes be demonstrated. For example, when tumour cells are injected into a histocompatible animal, a certain minimal number of cells (which varies with the particular tumour and with the age of the animal) is required to produce a tumour. This alone shows that the animal can destroy a limited (sub-threshold) number of tumour cells. Moreover, in an animal already bearing a tumour, the number of tumour cells of the same type which must be injected to produce a second tumour is often considerably greater than the threshold dose for a normal animal. This apparent paradox—that the animal can destroy an oncogenic dose of injected tumour cells while its original tumour continues to grow—has not been satisfactorily explained.

There is also evidence that animals can be actively immunised against tumours, e.g. by injecting (a) a sub-threshold dose of tumour cells, (b) a larger dose of tumour cells previously rendered incapable of dividing by radiation or cytotoxic drugs, or (c) membrane preparations of tumour cells. Animals so-treated are capable of rejecting a normally oncogenic dose of cells of the corresponding tumour. Protection against virus-induced tumours can also be provided by immunising against the virus. Only partial protection is provided by these procedures: it varies from tumour to tumour and can be overcome by injecting large numbers of tumour cells.

Experimental evidence indicates that immunological defence against solid tumours is attributable mainly to cell-mediated immunity and little, if at all, to antibody. For example, when lymphocytes from an animal which has been immunised to a tumour (as described above) are transferred to a second, syngeneic animal, they afford some protection against challenge with the same tumour: antibodies give little or no protection.

As already noted, antibodies are probably capable of killing free tumour cells in the blood and in exudates in the peritoneal cavity etc. This may result from the susceptibility of antibody-sensitised cells to the lytic effect of com-

plement, to phagocytosis and destruction by macrophages, and to the cytotoxic effect of K (antibody-dependent cytotoxic) cells (p. 152). Antibodies may thus play a role in preventing the formation of metastases by cancer cells gaining entry to the blood, serosal cavities, etc. However, as explained below, in some circumstances antibodies may actually enhance the growth of tumour cells.

The mechanism of destruction of tumour cells by specifically primed T-lymphocytes (Fig. 11.12) may be a direct cytotoxic effect requiring contact between the lymphocyte and target cell (p. 157) or may be mediated by lymphokines, including those which attract, immobilise and enhance the phagocytic and killing capacities of macrophages, and also by the specific arming of macrophages (p. 159).

Enhancement and 'blocking' factors. The injection into animals of antibody to the surface antigens of tumour cells has been observed, under certain conditions, to reduce the dose of the tumour cells necessary to cause a tumour, and to enhance the growth of a previously-implanted tumour. This **experimental enhancement of tumour growth** is antigen-specific; it is closely similar to the protection of tissue allografts by antibody (p. 167) and appears to be due to the antibody combining with antigen on the surface of the tumour cells and thus protecting them from attack by specifically primed lymphocytes. In other words, the antibody blocks the delayed hypersensitivity reaction of T lymphocytes with the tumour cells. In animals with large and progressing tumours, **blocking factors** have been detected in the serum; they have been shown to interfere with the killing of tumour cells *in vitro* by specifically-primed lymphocytes. It now seems unlikely that these are antibodies, for they disappear rapidly from the blood after excision of the tumour, at a time when the level of antibody increases. Other possibilities are that the blocking factors are free antigen molecules shed by the tumour cells as part of the normal turnover of plasma membrane constituents, or such antigens combined with antibody, i.e. immune complexes. Free antigen or immune complexes are capable of combining with the tumour-specific receptors on the surface of T lymphocytes or 'specifically armed' macrophages, thus preventing them from reacting with tumour cells: immune complexes might, in addition, bind to the Fc receptor sites of 'K'

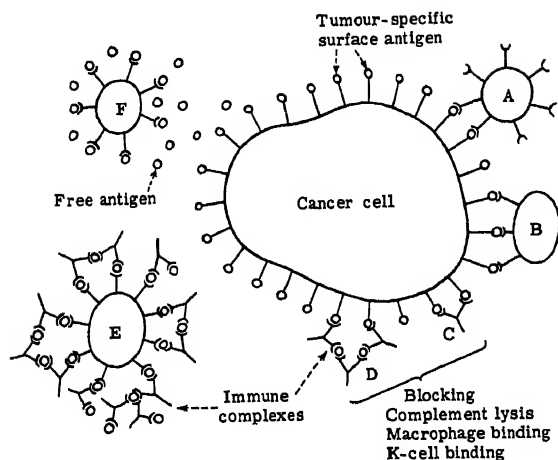


Fig. 11.12 Immunological cytotoxic mechanisms which may injure the cancer cell, and blocking factors which may protect the cell from immunological attack. The cancer cell may be killed by the reaction of cytotoxic T lymphocytes (A), or specially-armed macrophages (B), with tumour-specific surface antigen. Antibody (C) and antigen-antibody complexes in antibody excess (D) may protect the cancer cell from T lymphocytes and macrophages by binding to and blocking its surface antigens. Antibody and complexes may also promote cytotoxicity by complement, macrophages or K cells, but these effects seem to be of little or no importance except perhaps on free cancer cells in the circulation or in serosal exudates. Immune complexes in antigen excess (E) or free antigen (F) shed by cancer cells may block the specific binding sites on T lymphocytes or specifically-armed macrophages.

lymphocytes or macrophages, thus inhibiting their antibody-dependent cytotoxic activity for the tumour cells. These phenomena are shown diagrammatically in Fig. 11.12.

Immunology of cancer in man

Once a cancer has spread beyond the possibility of excision, it usually progresses, and eventually proves fatal. This applies to most but not all types of human cancer. There are, however, considerable variations in the rate of growth, even for tumours of the same type and histological appearances. In patients with carcinoma of the breast, for example, surgical excision is sometimes followed by many years of normal health, but with subsequent re-appearance and relatively rapid growth of tumour in the operation scar or of metastases: this is sometimes observed also with some other cancers. Very rarely, complete spontaneous regression of a cancer occurs. Partial regression is

more common, and it is not very unusual for patients with metastatic melanoma to have no obvious primary tumour: examination sometimes reveals an area of skin depigmentation which on histological examination shows evidence of a regressing malignant melanoma.

Observations of the sort outlined above suggest that defence mechanisms against cancer can develop in the host, and that in some instances these are partially (and rarely fully) effective. The nature of host defence in these instances is unknown, but there is circumstantial evidence that specific immune responses may influence the course of some tumours. In the rapidly growing form of breast cancer termed encephaloid cancer, for example, a favourable prognosis following excision has been reported to show some correlation with lymphocytic infiltration of the cancer, and this suggests that a delayed hypersensitivity reaction, or possibly antibody-dependent lymphocyte cytotoxic activity, is involved. Seminoma, a rapidly growing carcinoma originating in the germinal epithelium of the testis, is very commonly infiltrated with large numbers of lymphocytes and may also contain tubercle-like macrophage granulomas, both of which are consistent with a delayed hypersensitivity reaction. This may help to explain why the cure-rate is unusually high for such a rapidly progressing cancer: even when there is extensive metastasis, radiotherapy is often curative. A third example of a cancer which is often highly susceptible to therapy is choriocarcinoma arising from the placental trophoblast. Although it grows and spreads very rapidly, this rare tumour can often be destroyed by cytotoxic drug therapy. It is, however, unique among human tumours, in being a tumour of fetal tissue which grows in the mother; the tumour cells thus have HLA antigens which are inherited from the father and are foreign to the maternal host, so that allograft rejection is likely to contribute to the success of therapy.

Immune responses. *As in animal studies, antibodies and lymphocytes which react specifically with tumour cells have been demonstrated by in vitro techniques in patients with various types of cancer.* They react with the patient's own tumour cells, with the cells of other tumours of similar type, and with cell lines derived from such tumours. This sharing of common antigens by similar tumours in different individuals is consistent with, but not strong evidence of, a

viral aetiology. Antibodies and cell-mediated immunity to the tumour cells are most often detectable in patients with early cancer, and tend to diminish as the tumour enlarges and spreads. Blocking factors, similar to those in animals (see above) also appear in the blood, particularly when the cancer is advanced. These observations are based on recent work, and the prognostic significance of fluctuations in the immune responses is not yet known. In a study of neuroblastoma, a malignant tumour of infancy and early childhood, which regresses completely much more often than most other cancers, Dr. Lindsay Morrison, working in Glasgow, has demonstrated the development in patients of antibody and cell-mediated immunity to the tumour, but in this instance the immune response does not appear to correlate with spontaneous regression. The antigenic preparations in such investigations have consisted of whole tumour cells or crude homogenates or extracts, and the tumour-specific cell-surface antigens have not yet been fully characterised. Like HLA antigens, they have been reported to contain a β_2 -microglobulin chain.

Oncofetal antigens are so called because they were first detected as products of cancers and of fetal tissues. Sensitive radio-immunoassay techniques have, however, detected low levels of them in normal adult serum and tissues and in raised amounts in some patients with cancer or with non-neoplastic cellular proliferation. The serum level of *carcino-embryonic antigen* (CEA) is raised in many patients with colorectal or various other cancers and in some patients with hepatitis, chronic bronchitis or ulcerative colitis etc. Raised levels of α -fetoprotein (AFP) are *relatively* specific for liver-cell cancer and malignant teratoma but occur also in some patients with other forms of cancer or hepatitis etc. Alpha-fetoprotein assay is of value in the ante-natal detection of neural tube defects (p. 772).

Immunological surveillance

The concept of immunological surveillance was advanced by Burnet, who postulated that one function of T lymphocytes is to monitor host cells and respond immunologically against any which have developed surface antigens foreign to the host. By this means, it is conceivable that many (perhaps most) malignant cells are destroyed before they are capable

of developing into a tumour. As regards **experimentally-induced tumours**, there is little evidence that this hypothesis applies to those caused by chemical carcinogens, for such tumours are produced no more readily in immunosuppressed animals, or in congenitally immunodeficient (nu nu) mice than in normal animals (Stutman, 1979). Immunological surveillance is, however of considerable importance in preventing virus-induced tumours in animals. This is illustrated by the ease with which such tumours can be induced in neonatal mice, which are immunologically immature, and in adult immunosuppressed or immunodeficient mice. Protection of neonates or immunodeficient mice is also afforded by transfer of syngeneic lymphocytes from an animal immunised against the virus-induced tumour. Indeed, immunosuppression by the oncogenic virus itself, as in the case of feline leukaemia virus, predisposes to tumour production.

As regards **cancer in man**, there is also some evidence favouring immunological surveillance, for children with certain congenital immunodeficiencies, and renal transplant recipients receiving long-term immunosuppressive therapy, have an increased incidence of cancer. However, the tumours arising in such patients do not reflect the natural incidence of cancer in the general population. The increased incidence is greatest for tumours of the lymphoreticular system (lymphomas and lymphoid leukaemias). These include tumours similar to those induced in animals by the leukaemia viruses, and it thus seems likely that such tumours in man, as in animals, are susceptible to immunological surveillance.

The most likely example of effective immunological surveillance in man is provided by infection with the Epstein-Barr virus (p. 304) which in normal individuals induces a strong immune response on the part of T cells, and does not cause cancer. In African children with chronic malaria, which depresses the immunity system, infection with EB virus is associated with the development of the Burkitt lymphoma, and in the Southern Chinese, in whom genetically-determined immunological responsiveness may play a role, the virus is associated with nasopharyngeal carcinoma.

In the past few years, there has been increasing interest in the possibility that immunological surveillance is a function of '**natural killer**' (NK) **lymphocytes**, which are present

in the blood of normal animals, including man. NK cells can react with and kill foreign cells, including tumour cells, apparently without the need for previous immunological priming (see Herberman and Holden, 1978). The nature and importance of NK cells are not yet established. It has been claimed recently that their reaction with foreign cells depends on surface antibody bound to the NK cell by its Fc receptors (Takasugi and Akira, 1979). If this is correct, it is not clear how NK cells differ from K cells.

It may be significant that most cancers occur in old age, for there is no doubt that immune responsiveness declines in the elderly, but prolonged exposure to environmental carcinogenic factors and the long latent period of human cancers could also account for the age incidence. Tests for immune responsiveness to various antigens have not, in general, revealed immunodepression in patients with early cancer as compared with age-matched control subjects. Advanced cancer patients commonly show evidence of immunodepression, but this is most obvious in those with tumours which invade and destroy the lymphoid tissues, and is likely to result from the cancer.

The prospect of immunotherapy. The possibility of immunotherapy for cancer has been appreciated since the early 1900s. The advances in tumour immunology outlined above have strengthened the scientific basis of such treatment, but they have also demonstrated that, in spite of the occurrence of anticancer immune responses in many patients, their tumours still progress and cause death. It remains possible, however, that boosting the immune response might have some therapeutic effect. The administration of immunological adjuvants, such as BCG, *Corynebacterium parvum* and Levamisole, has been attempted in various neoplastic conditions, and has been claimed to have some effect in acute lymphoblastic leukaemia: in other neoplastic conditions the results have so far been disappointing. Various attempts have also been made to stimulate active specific immunity by implanting pieces of tumour which have been excised and treated with x-irradiation to prevent the cells from dividing. Such a procedure is not very hopeful, for if the patient's tumour does not stimulate effective immunity it seems unlikely that the implanted cells will do so, but it remains possible that, by increasing the antigenicity of the implanted cells, e.g. by coupling with haptens, or by using

homologous tumour with its 'foreign' transplant antigens, the immune response to the relevant tumour antigens may be augmented. Attempts have also been made to provide passive immunity by transplanting tumour to a volunteer in the hope that therapeutically effective antibody will be produced. On at least one occasion, the volunteer failed to reject the transplanted tumour, which proved to be fatal. The therapeutic value of lymphocyte products, and of interferon (p. 194), are being investigated.

Immunotherapy of cancer patients faces at least three major difficulties. Firstly, excision of an early cancer may effect a cure. It is not possible, at present, to identify those patients who will develop recurrences, and it therefore seems unjustifiable to apply to early cancer patients a form of therapy which is of unknown value. Accordingly, attempts at immunotherapy have mostly been made on patients with advanced cancer, when it is likely to be too late. Secondly, there is no guarantee that active immunisation will induce immune responses which contribute to the destruction of the tumour. There is, in fact, a risk of inducing the production of 'enhancing' antibody (p. 315) and thus increasing the rate of tumour growth. Thirdly, problems in assessing the results of any form of cancer therapy arise from the natural individual variations in the rate of growth and spread of cancer. In consequence, any trial of therapy must usually be extensive.

Attention has also been given to the possibility of vaccines to prevent cancer, and this has been achieved in feline leukaemia (p. 307). If it could be shown that some forms of human cancer are due to particular oncogenic viruses, immunisation against the virus, or against tumours induced by it, should be possible, but it would still be necessary to identify those individuals likely to develop that form of cancer unless one is prepared to immunise whole populations.

Finally, the induction of a delayed hypersensitivity reaction at the site of a tumour has been used as a method of tumour destruction. This has achieved some success in the treatment of epidermal tumours, notably basal cell carcinoma. The patient is sensitised by application to the skin of an agent which induces cell-mediated immunity, e.g. dinitrochlorobenzene (DNCB) and subsequently DNCB is applied to the tumour and surrounding skin: a delayed

hypersensitivity reaction develops, and may be successful in destroying the tumour. Two mechanisms may be involved: firstly, the delayed hypersensitivity reaction, if intense, causes necrosis of normal cells (as in tuberculin skin testing—p. 157) and may similarly induce necrosis of the tumour. Secondly, macrophages accumulate and become more actively phagocytic and

cytotoxic for foreign cells (including cancer cells) unrelated to the antigen which has induced the delayed hypersensitivity reaction. Similarly, attempts have been made to destroy tumours by immunising the patient with BCG (p. 209) and injecting the tumour with either BCG or tuberculo-protein.

What makes the cancer cell multiply?

Although we are aware that many agents can cause individual cancers, we know neither how they induce the essential change in the cell which makes it a cancer cell, nor indeed what is the exact nature of this change. This last section will review briefly the major theories which attempt to account for the most characteristic feature of the cancer cell—its property of multiplying regardless of the mechanisms which govern the behaviour of the normal cell.

Cell surface changes. The inherent nature of the cancer cell abnormality is reflected in its loss of contact inhibition when grown in culture (p. 302), a phenomenon which points to an abnormality of the cell membrane.

It has been shown that most types of tissue cell form *gap junctions*, through which there is continuity of the cell sap of adjacent cells. These connections may be of importance in transmitting signals from cell to cell, for cancer cells in general lack gap junctions. The surface of cancer cells also carries a higher negative charge than most normal tissue cells, and increased mutual repulsion may thus interfere with the adhesion and contact inhibition of cancer cells.

Two other cell membrane components of possible importance in the mitotic activity of cancer cells are proteins and sugar residues, e.g. N-acetylglucosamine, on the cell surface. It is known that mitosis is associated with some loss of surface protein, and normal tissue cells in culture can be induced to multiply by treatment with trypsin: after mitosis, however, the daughter cells develop surface proteins and contact inhibition is restored. By treating cancer cells in culture with compounds which bind to surface N-acetylglucosamine, contact inhibition can be restored and the cells cease to multiply. It is postulated that exposed sugar residues on the

cell surface are involved in the signal which stimulates cells to multiply, and that surface protein in some way inhibits this role of the sugar residues.

The signal at the cell surface appears to be relayed within the cell by its effect on intracellular cAMP, the level of which falls during mitosis, and cGMP, which rises. Malignant cells have, in general, low concentrations of intracellular cAMP, addition of which to transformed cells in culture can inhibit their growth.

A defective response to *chalones* (p. 86) has also been postulated to account for proliferation of cancer cells, but is at present little more than speculative.

Whatever the nature of the essential change in the cancer cell, it is obviously transmitted to the daughter cells during mitosis, and this raises two possibilities. Firstly, that carcinogenesis involves *mutations*, i.e. abnormalities in the genome of the cell, and secondly that the cancer cell represents *abnormality or reversal of differentiation (the epigenetic theory)*.

The mutational theory of carcinogenesis

This proposes that the altered behaviour of the cancer cell arises from mutation. In support of this theory, most chemical and physical carcinogenic agents are mutagenic, chromosomal anomalies are a common feature of cancer, and oncogenic viruses bring about their effects by the activity of a viral oncogene integrated into the cell DNA, which itself could be regarded as the equivalent of a mutation. However, we have seen that carcinogenesis is usually a gradual process. With the exception of some of the oncogenic viruses, carcinogenic agents bring about gradual changes in the cell, illustrated above by the effect of azo-dyes on liver cells (p.

300). If mutation is the essential change in carcinogenesis, then it is necessary to postulate a series of mutations.

Related to the mutation theory is the *oncogene theory*. As already noted, the proviruses of endogenous oncogenic reproviruses (p. 309) are demonstrable in the normal cells of various vertebrates. In such cells, the oncogene of the provirus is inactive: the oncogene theory proposes that carcinogenic agents bring about cancerous transformation by inducing changes which result in expression of the proviral oncogene. This might conceivably occur without expression of other proviral genes, in which case there would be little or no evidence of the role played by the endogenous virus. If, however, the virogene is fully expressed, then viral products are likely to be detectable in the cancer cell, and unless the provirus is deficient (p. 306), the development of cancer would be associated with viral replication, as occurs with some of the leukaemia viruses.

The epigenetic theory of carcinogenesis

Although the mutation theory of cancer is widely favoured, it must be emphasised that heritable changes in cells occur without mutations. With the exception of committed lymphocytes and their progeny (pp. 126–9), all somatic cells are believed to possess the whole genome of the individual, and this has been supported by the growth of plants from single cells and the development of a normal frog when the nucleus of a fertilised frog ovum is replaced by a frog's somatic cell nucleus. Nevertheless, during differentiation, the somatic cells develop special features which characterise them as neurons, liver cells, fibrocytes, etc., and this differentiation is retained by its descendants when the differentiated cell divides. It is thus conceivable that the special features of the cancer cell have developed as a result of reversal of the process of differentiation, or of abnormal differentiation of a primitive stem cell. While at first sight the cancer cell might appear to have gained positive properties—increased mitotic activity, motility, invasiveness, production of inappropriate and excessive amounts of hormones, etc.—these may, in fact, represent *loss of cell components and consequent failure of normal homeostasis*. As noted above, uncontrolled growth of cancer cells appears to be

closely associated with loss of a surface protein. Similarly, production of excess of hormones or of inappropriate hormones could be due to loss of the suppressor mechanism for a particular gene. Such loss of controlling factors could be explained by a particular pattern of gene expression, i.e. a form of differentiation, without the need to postulate mutations.

A number of phenomena support the epigenetic theory of cancer (Uriel, 1979). For example, some undoubtedly malignant tumours stop growing and their cells become highly differentiated. A good example is the neuroblastoma, whose cells may develop into mature neurons. This is a rare happening, but even in the common squamous carcinoma many of the tumour cells stop dividing and become highly keratinised (Fig. 12.13, p. 329). Malignant cells in culture can also, in some instances, be made to regain contact inhibition and other features of normal cells by addition to the culture medium of various chemicals, e.g. 5-bromodeoxyuridine. Of particular interest is the demonstration that replacement of the nucleus of a fertilised frog's ovum (see above) by the nucleus of cells from a Lucké carcinoma (a virus-induced renal cancer of the frog kidney), does not result in the growth of a tumour cell-line but of a normal tadpole. This implies that the normal fertilised ovum and its progeny contain cytoplasmic factors which can suppress the activity of an integrated viral oncogene, and that the cancer cell lacks such factors. Other important findings have arisen from the implantation of malignant cells from mouse gonadal teratomas (tumours derived from primordial germ cells of the mouse testis or from parthenogenetically activated ova) into mouse blastocyst embryos. The malignant cells divide during embryonic development, producing cells which differentiate normally and contribute to the formation of several tissues. By contrast, when the teratoma cells are implanted into adult mice, they produce a teratoma. It is thus apparent that, in this instance, the environment of the embryo induces normal behaviour in the malignant cells.

While of limited scope, experiments such as those described above indicate the need for further studies on the relative importance of nuclear and extra-nuclear changes in carcinogenesis.

In conclusion

The essential nature of carcinogenesis—mutations, epigenetic changes or oncogenes—has not been settled. It may well be that more than one type of fundamental cellular change can result in cancer. The protective host factors are really no better understood: immunological surveillance appears to play an important role in some of the human lymphomas, and by analogy with animal studies this suggests a viral aetiology for these tumours.

So far, experimental studies on the fundamental nature of cancer have contributed little to cancer prevention in man, but detection and elimination of exposure to chemical and physical carcinogenic agents has without doubt achieved considerable success. The curtailment

of such habits as cigarette smoking and betel chewing could also prevent huge numbers of cancers.

In many countries, cytological screening for uterine cervical pre-malignancy appears to have reduced the incidence of invasive cancer, and the earlier detection of cancer of the breast and some other sites is being attempted by education and screening of the public.

Once cancer has developed and spread far beyond the possibility of complete removal, symptomatic relief of pain may be offered by radiotherapy and cytotoxic drugs, or in some instances by hormonal therapy; such therapy may slow down or even for a while arrest the growth of the tumour. Complete cure of advanced malignancy is, however, a rarity except in a few particular types of cancer.

References

- Herberman, R. B. and Holden, H. T. (1978). Natural cell-mediated immunity. *Advances in Cancer Research* **27**, 305–77.
- Hill, M. J. (1977). Bacterial Metabolism pp. 45–64. In *Topics in Gastroenterology*. Ed. by S. C. Truelove and E. Lee. Blackwell Scientific, Oxford.
- Stutman, O. (1979). Chemical carcinogenesis in nude mice from heterozygous matings and homozygous matings. *Journal of the National Cancer Institute* **62**, 353–8.
- Takasugi, M. and Akira, Donna (1979). Role of antibodies in specificity of natural cell-mediated immunity. *Journal of the National Cancer Institute* **62**, 1361–5.
- Uriel, J. (1979). Retrodifferentiation and the fetal patterns of gene expression in cancer. *Advances in Cancer Research* **29**, 127–74.

Further Reading

- Advances in Cancer Research*. Academic Press Inc., New York. (Detailed reviews of oncological topics of major importance—mostly excellent. Usually one volume published annually since 1953.)
- Cochran, A. J. (1978). *Man, Cancer and Immunity*, pp. 206. Academic Press, London, New York and San Francisco. (A review of immunological aspects of cancer, including some of the author's recent work.)
- Neville, A. Munro, Grigor, K. M. and Heyderman, Eadie (1978). Biological markers and human neoplasia. In *Recent Advances in Histopathology*, No. 10, pp. 23–44. Edited by P. P. Anthony and N. Woolf. Churchill-Livingstone, Edinburgh, London and New York.
- Symington, T. and Carter, R. L. (Eds.) (1976). *Scientific Foundations of Oncology*, pp. 690. Heinemann, London. (Authoritative reviews on many aspects of oncology.)
- Taussig, M. J. (1979). Neoplasia. In *Processes in Pathology*, pp. 241–353. Blackwell Scientific, Oxford. (A more detailed account than provided in this chapter—clearly written and well illustrated.)

Tumours: II. Epithelial Varieties and Modes of Spread

After the introductory account of tumours in the last chapter, we turn for the next two chapters to the more practical questions of what kinds of tumour occur in man, what they look like and how they behave. There are many aspects of tumours that are best described as part of the pathology of the organs from which they are derived: these chapters are concerned with aspects of more general application, though they will be found to include also descriptions of a number of specialised tumours, usually because they illustrate some general principle.

Classification

In the introduction to the previous chapter the distinction between *benign* and *malignant* tumours was discussed in some detail. We must now use the other chief mode of classification, the *histogenetic*, which is based on the tissue of origin. We may conveniently distinguish tumours of:

- (a) epithelium;
- (b) connective tissues (including muscle);
- (c) blood vessels and lymphatics;
- (d) the nervous system;
- (e) lymphoid and haemopoietic tissue;
- (f) other tissues.

Of these, the epithelial tumours are overwhelmingly the commonest and are responsible for 90 per cent of all cancer deaths in this country. This first chapter will be devoted to them alone.

General features of epithelial tumours

Epithelium has two essential characteristics which are carried over into its tumours.

- (a) It forms continuous sheets or masses of

cells of similar type, which adhere together without any intervening intercellular structures (Fig. 12.1). This adherence of the cells into larger or smaller groups is retained as an indication of an epithelial origin even in tumours which have lost all the other distinctive features of epithelial cells.

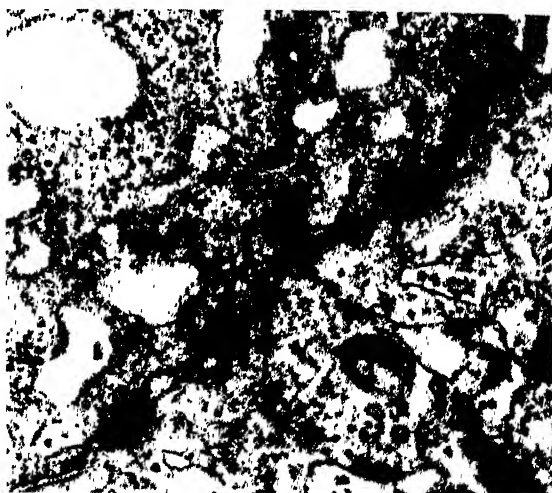


Fig. 12.1 The junction between two epithelial cells showing their close apposition. Two zones of adherence (desmosomes) are seen as double lines of increased density of the cell membranes. Desmosomes are observed in the cells of malignant epithelial tumours, but are reduced in number or defective with consequent weakening of cell adhesion. $\times 40\,000$.

- (b) It requires a stroma of connective tissue and blood vessels for its support and nourishment. This is equally necessary for tumour epithelium, and all epithelial tumours appear to be able to stimulate the local connective tissues and blood vessels to proliferate and supply a stroma which surrounds and supports the epithelial cell groups. This 'desmoplastic reaction', as it is called, varies in degree; it is often inadequate, so that much of the tumour dies from ischaemia, but it is sometimes excessive,

so that the fibrous stroma becomes more conspicuous than the epithelium ('scirrhous' tumours, so-called). We know very little about the way in which the tumour cells influence the development of connective tissue. As so often in cancer studies, the basic problem is a much wider one: the relationship between epithelium and connective tissue is established as a convenient form of organisation in a large part of the animal kingdom, and we know little of its basic

mechanism. The cancer cells are simply exploiting a normal process.

The way in which an epithelium is organised naturally affects profoundly the structure of the tumours to which it gives rise, and this is especially marked in the case of the slow growing and well-differentiated benign tumours. Epithelia which cover surfaces generally give rise to **papillomas**; epithelia of exocrine or endocrine glands, and of solid organs like the liver and kidneys, give rise to **adenomas**.

Benign Epithelial Tumours (Papillomas and Adenomas)

Papillomas

If one considers what will happen to a sheet of epithelium such as the epidermis when its cells have begun to multiply, it is clear that the first effect will be to thicken the layer. But since this is limited by the extent to which nutriment can diffuse from the underlying blood vessels, the tumour cells must soon spread in other directions. So long as the tumour is benign, they do not spread downwards into the underlying tissue, and so the epithelium must increase in thickness or in surface area. Most often it increases in both. The effect of increase in area is most readily understood if one tries to visualise the epithelium as a sheet of cloth which is pinned down at the edges and then increased in area; it is obvious that it will be thrown into folds. If the increase is in one dimension only, the folds will be regular pleats, but since in a benign epithelial tumour it occurs in two dimensions simple folding cannot occur, and the result is an irregular mass of peaks and hollows. Where the epithelium is raised into peaks, the underlying connective tissue proliferates and accompanies it, forming a fibrous core (the 'desmoplastic reaction' in effect) and the epithelium covering it remains well nourished and continues to grow. Since each peak is compressed by the other peaks around it, it can only grow upward, producing a higher peak which may finally become a long finger-like process, or *frond*. The resulting mass of 'papillae' constitutes a papilloma.

If such a papilloma arises in a squamous epithelium the processes are naturally covered by squamous epithelium, thickened but other-

wise not grossly abnormal. They are well seen in a papilloma of the skin (Fig. 12.2), the commonest form of which is the virus-induced wart of children.



Fig. 12.2 Papilloma of muco-cutaneous junction of lip, showing branching processes of connective tissue covered by stratified epithelium. $\times 10$.

The transitional epithelium of the urinary tract produces papillomas covered by transitional epithelium, in which the papillary processes are often very numerous, long and thin: such tumours are sometimes called **villous**

papillomas.* (Figs. 12.3 and 12.4). This is possibly due, not to any special characteristic of the epithelium, but to the environment provided by the bladder, in which the fronds of the papilloma float like seaweed in a sheltered bay: similar complexity is seen in the rare choroid plexus papillomas which float in the cerebrospinal fluid.



Fig. 12.3 Papilloma of bladder, showing innumerable delicate fronds, which are covered by transitional epithelium. A stalk, as seen in this case, is often absent. It is unusual for a bladder papilloma as large as this to be completely benign. $\times 4$.

Epithelium lining ducts, e.g. of the breast, can give rise to papillomas (Fig. 12.5) and villous papillary tumours arise also from the surface epithelium of the large intestine (Fig. 12.6), although here it is commoner to find more complex tubulovillous tumours (described on p. 327).

The papillary structure of papillomas is not usually obvious to the naked eye, though the educated naked eye may recognise it. In most skin papillomas it is hidden by the thick horny layer of keratin which develops on the surface, filling up the gaps between the fronds and producing a rough dry hard surface in which only an ill-defined cauliflower pattern gives a hint of the underlying structure. The result is a well defined little lump, usually round, always projecting above the surface (except for the plantar wart on the sole of the foot, where pressure flattens it) and at times slightly polypoid—i.e. having a slight neck between it and the skin level. The villi of excised bladder papillomas

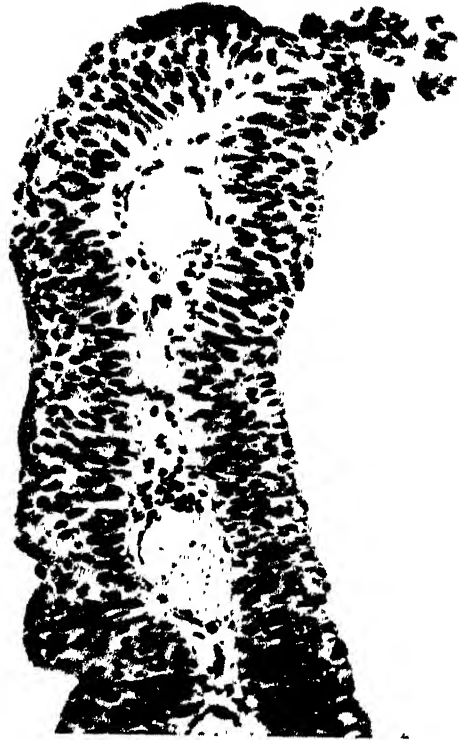


Fig. 12.4 Papilloma of bladder. Section of the tip of a single process, showing a narrow core of connective tissue, including capillaries, covered by transitional epithelium. The epithelium is rather thicker and has rather more cells than normal, but differentiation is still excellent. $\times 200$.

are difficult to recognise by naked-eye; unless the excised tumour is submerged in fluid, they collapse against each other and leave nothing but a somewhat velvety surface to indicate their true nature. Seen *in situ* with a cystoscope, or examined with a lens under saline, the fronds will be obvious. The fronds are even harder to see in the soft velvety plaques of a villous tumour of the rectum, and a dissecting microscope may be necessary.

Adenomas

Adenomas are benign epithelial tumours of glandular origin and often retain some secretory function. They occur more commonly, in general, in the endocrine than in the exocrine glands, though the commonest of all are the tubular adenomas (adenomatous polyps) arising from the mucosa of the large intestine.

* Villi: a fine hair-like process.



Fig. 12.5 Papilloma of the breast, growing into and distending a duct near the nipple. $\times 4$.



Fig. 12.6 Papillary tumour of rectum. The structure is obviously similar to that of the bladder papilloma, but a higher power would show that the covering epithelium is of columnar type. The majority of benign tumours of colorectal epithelium are of mixed papillomatous and adenomatous structure and are rather misleadingly termed adenomas. $\times 3$.

Adenomas of endocrine glands

These are of particular importance because many are capable of excessive hormonal secretion. This is a particularly common feature of adenomas of the parathyroid, the islets of Lan-

gerhans and the adrenal medulla, all of which usually reveal their presence by excessive secretion of one or more of the appropriate hormones of the parent gland. Some adenomas of the adenohypophysis, thyroid and adrenal cortex behave in this way, although most such tumours do not function sufficiently to cause hormone imbalance. As with tumour cell function in general, the secretory activity of most endocrine adenomas is not subject to control by the normal feedback mechanisms.

In all these glandular tissues an adenoma usually appears as a rounded or lobulated nodule, generally solid and of the same colour as, or paler than, the surrounding tissue (Fig. 12.7). It is enclosed in a fibrous capsule, usually thin, and resulting from pressure atrophy of the surrounding glandular tissue and condensation of its stroma (Fig. 12.8). Adenomas range in size from the microscopic to over 10 cm in diameter, but most of those of the endocrine glands, with the exception of thyroid adenomas, are less than 2 cm in diameter.

As might be expected from their functional activities, the cells of endocrine adenomas re-



Fig. 12.7 An adenoma of the thyroid gland. The tumour is partly enclosed in normal thyroid tissue and is enclosed in a fibrous capsule, most clearly seen around the lower margin. The cut surface of the tumour resembles thyroid tissue. $\times 1.5$.



Fig. 12.8 Part of a thyroid adenoma of the 'solid' or micro-acinar type, which contrasts in appearance with normal thyroid tissue. Note the fibrous 'capsule' which is composed largely of residual stroma of compressed, atrophic thyroid surrounding the tumour. $\times 40$.

semble closely those of their parent tissue; adenomas of those endocrine glands composed of several cell types, e.g. the adenohypophysis and pancreatic islets, may be predominantly of one particular cell type and secrete the particular hormone(s) normally secreted by that type of cell. For example, islet cell adenomas may be composed predominantly of β -cells and secrete insulin, of α -cells and secrete glucagon, or of δ -cells and secrete somatostatin.

Although endocrine adenomas are usually well differentiated, they often show great variation in nuclear and cell size, which in these tumours does not suggest malignancy unless accompanied by other features (numerous mitoses, invasiveness, etc.). The stroma may also resemble that of the parent gland, although it is often more dense, abundant and sometimes hyaline. Illustrations of endocrine adenomas can be found in Chapter 26.

Adenomas of the exocrine glands

These are uncommon, apart from the very common **fibro-adenoma of the breast** (Fig. 24.41), which is really a mixed tumour in which both the epithelium and stroma are neoplastic. Both the prostate and breast are very prone to develop multiple nodularity due to foci of proliferation of both glandular and stromal elements in various proportions, but such nodules are not sharply defined and are really examples of focal hyperplasia, perhaps due to hormonal influences, and are not true tumours. Adenomas of the **salivary glands** are not rare. They show pleomorphism of both epithelial and stromal elements (Figs. 19.14, 19.15, p. 596) and projections through the capsule make their complete removal more difficult.

Cystadenomas. Curiously, the **ovary** is a common site of an unusual type of adenoma in which the cells arrange themselves in gland-like structures resembling acini. The cells secrete copious mucous or watery secretion and as they have no ducts into which this can drain, it



Fig. 12.9 Mucinous cystadenoma (cystic adenoma) of ovary. The cysts (small in this case) are lined by tall mucin-secreting epithelium. The nuclei, situated at the base of each cell, form a continuous line hardly distinguishable in this picture from those of the next cyst. $\times 150$.

distends the lumen of the acinar-like structures, resulting in large cystic spaces and giving the tumour its name, **cystadenoma**. Epithelial-lined processes may project into the lumens (**papillary cystadenoma**). Such tumours, particularly those which secrete mucin (Fig. 12.9 and Fig. 24.20, p. 961), may grow enormous, and examples are on record of such tumours which outweighed the patient!

It might be expected that the pure cystadenomas, in which secretion is abundant and hence differentiation is good, are more benign than papillary cystadenomas, in which proliferation is more active: this is generally, although not always, true.

Papillary cystadenomas occur rarely in the pancreas and kidney.

Colorectal adenomas. The common **tubular adenomas of the large intestine** are composed of mucus-secreting cells arranged in tubular glands, thus resembling in structure the normal mucosal glands, and they are surrounded by a stroma

similar to the lamina propria. These tumours are not embedded in the mucosa and are not encapsulated. Each consists of a rounded or lobulated mass which projects from the mucosal surface, and as it enlarges it becomes pedunculated, i.e. develops a stalk or pedicle, with a fibrovascular core and lined by normal mucosa, from which it dangles in the lumen of the bowel. Such a tumour is termed an **adenomatous polyp** * or **tubular adenoma** (Fig. 12.10). Although the glands secrete mucin, most of them drain (like the tubular glands of normal mucosa) into the lumen of the bowel and so do not become distended. As already mentioned (p. 324) papillary tumours ('villous adenomas') also arise from the large intestinal mucosa, and many epithelial tumours of this site have a complex structure in which papillary and adenomatous elements are mixed (tubulo-villous adenomas).



Fig. 12.10 Tubular adenoma of the colon. The rounded darker mass of the adenoma is made up of close-packed glands, less regular and more cellular than those of the normal mucosa which is seen below, covering the stalk of the polyp. (Unusually, two smaller adenomas arise from the stalk. In this case there were multiple adenomas and a carcinoma, seen in Fig. 12.14.) $\times 8$.



Fig. 12.11 Part of tubular adenoma of the colon, showing cellular aberration. Although the epithelium still forms crypts, these are hypercellular and irregular in size and shape. The epithelial cells show abnormal proliferative activity which has resulted in their becoming tall and narrow, and in some the nucleus has left the basal position and is undergoing mitosis. Many of the cells no longer contain the large globule of mucin characteristic of goblet cells. $\times 150$.

*A polyp is a lump of tissue at the end of a stalk. It is not necessarily neoplastic.

Another important feature of these large intestinal tumours is that the glandular epithelium often shows marked mitotic activity and the cells become squeezed together and hence tall and narrow (Fig. 12.11): they may also become heaped up into two or more layers, or even form solid groups. Cell aberrations, such as nuclear irregularity and enlargement and cytoplasmic basophilia, are also common, and this proliferative activity and cell aberration is a warning that many of these tumours are pre-malignant: they are prone to invade the stalk and adjacent bowel wall and spread by lym-

phatics and blood stream. In patients with large numbers of adenomatous polyps (*polyposis coli*), the risk of malignancy is very high.

It is noteworthy that adenomas of exocrine glands do not show the high degree of specialisation and functional activity characteristic of many endocrine adenomas: they secrete a thin watery or mucoid fluid, but not the various enzymes secreted by the exocrine glands.

Adenomas occur also in the cortex of the **kidney** and rarely in the **liver**, the latter being of interest because their incidence appears to be increased in women using oral contraceptives.

Malignant Epithelial Tumours (Carcinomas)

The term **carcinoma** may be applied to any malignant tumour of epithelial origin. It may arise from one of the benign epithelial tumours just described, or directly from a non-neoplastic epithelium. In either case, it retains the two features already described as characteristic of epithelium and its tumours—the formation of sheets or masses of contiguous tumour cells, and the ability to excite a stromal reaction between and around the tumour-cell masses. In addition, the epithelial element commonly retains some resemblance to the tissue of origin, though in poorly differentiated (and usually more malignant) tumours this may be tenuous.

Naked-eye appearances

While there are many variations, it is possible to describe a typical carcinoma. It forms a firm lump, often irregularly nodular, its edge well defined in places and in others blending into the surrounding tissue (areas of invasion) so that it cannot be dissected out cleanly. On section it is predominantly whitish, as are most dense collections of young cells: there are often red patches of haemorrhage and, especially towards the centre, yellow areas of necrosis. When carcinoma arises in the epidermis or other surface epithelium it forms at first an irregularly dome-shaped swelling. The centre of this swelling however has often a poor blood supply and has lost the surface epithelium: it is exposed to trauma, infection and (in the case of lesions in the gut) digestive juices. It therefore often sloughs out, leaving a ragged ulcer. At

the edges the tumour has a better blood supply and is partly protected by the surface epithelium, and so survives: the ulcer therefore commonly retains a thick irregular raised edge which is responsible for the highly characteristic appearance of the ulcerated malignant tumour (compare, for instance, Figs. 12.14, 19.36, p. 615 and 27.36, p. 1075).

Varieties of carcinoma

There are many special types of carcinoma characteristic of particular sites, though the more highly malignant 'anaplastic' ones all tend to look alike. Most malignant epithelial tumours can, however, be included in the two great classes of **squamous carcinoma** and **adenocarcinoma**.

Squamous carcinoma

This is the characteristic malignant tumour of squamous epithelia, both epidermis and squamous mucosae. (The names *squamous-cell* and *epidermoid* carcinoma are sometimes used instead of squamous carcinoma, but have no particular advantages). In addition there are some unexpected sites, where squamous carcinomas arise in organs containing no squamous epithelium: the most important of these is the bronchus, where squamous metaplasia of the bronchial epithelium is the probable explanation; similar metaplasia can account for less common sites such as the urinary tract and the gall bladder. Unstable squamo-columnar

junctions such as the uterine cervix are also important sites.

Histologically, most squamous carcinomas are very readily recognisable. In early lesions, the downgrowth of the surface epithelium into the deeper tissues can be detected (Fig. 12.12). At the



Fig. 12.12 Squamous carcinoma of the tongue, showing squamous cell masses with, in places, central keratinisation. These remain in continuity with the over-lying epithelium (*above*). Ulceration is beginning (*above left*). There is a well-marked inflammatory infiltrate in the connective tissue, which a little obscures the distinction from the infiltrating tumour. $\times 38$.

periphery, small masses and narrow columns of cells burrow into the surrounding tissues. Behind this margin, the invading cell groups have had time to enlarge and differentiate, becoming recognisable as prickle cells and usually forming keratin: the keratin forms rounded concentric nodules in the centre of the cell groups, a very characteristic appearance called 'cell nests' or 'epithelial pearls' (Fig. 12.13). The amount of keratin formed and the proportion of cells recognisable as prickle cells both vary greatly: they are the best guides to degree of differentiation of the tumour, which, of course, affects its prognosis.

There are some variations from site to site—



Fig. 12.13 Squamous carcinoma at higher magnification, showing cell nests. The largest shows central keratin (still with some stratum granulosum granules), then large pale prickle-cells and a periphery of darker undifferentiated cells. In the smaller cell masses at the top keratinisation has not yet begun: the mass at left centre is at an intermediate stage. $\times 290$.

for instance, a squamous carcinoma of the bronchus or pharynx is usually less well differentiated than one of the lip or the skin. The site also naturally has a profound effect on the signs and symptoms produced by these tumours, and on their accessibility for treatment and hence on their prognosis.

Adenocarcinoma

This second group is a little less homogeneous than the last. Thus a histological section of an adenocarcinoma of the stomach can usually be distinguished from one of the colon with more confidence than a squamous carcinoma of the tongue from one of the bronchus. The grouping of adenocarcinomas together is, however, useful, for most of the malignant tumours of glands have a great deal in common

with each other and with those that arise from all the ducts and surfaces lined by columnar epithelium. Important sites of origin include the stomach and colon, the pancreas, gall bladder and its ducts, breast and uterus; also the bronchi, which can give rise to both squamous and adeno-carcinomas.

Histologically, almost everything that has been said of the adenomas applies to adenocarcinomas, with two differences.

(a) Instead of remaining localised, the tumour cells invade the surrounding tissues (Fig. 12.14).



Fig. 12.14 Adenocarcinoma of colon. Ulcerated tumour to left, normal mucosa to right: between them the raised 'rolled margin' formed by a thick layer of tumour still partly protected by the normal mucosa stretched over its upper surface. Invasion of submucosa and muscularis is well seen. $\times 11$.

(b) Differentiation is poorer (Fig. 12.15). In addition to all the general features of malignant tumours listed in the last chapter (p. 293) there is a marked tendency for acini to contain less secretion, to be lined not by one regular layer of epithelial cells but by a thick irregular layer, and in some tumours for most of the cell groups to form solid masses.

Several variations upon the basic pattern are common enough to be worth describing. Mixed and intermediate forms occur, and none of the following should be regarded as completely distinct entities.

(a) '**Spheroidal-cell carcinoma**', is a name commonly used for an adenocarcinoma in which most of the cell masses are solid (Fig. 12.16). This type of tumour is common in the breast, partly because of relatively poor differentia-

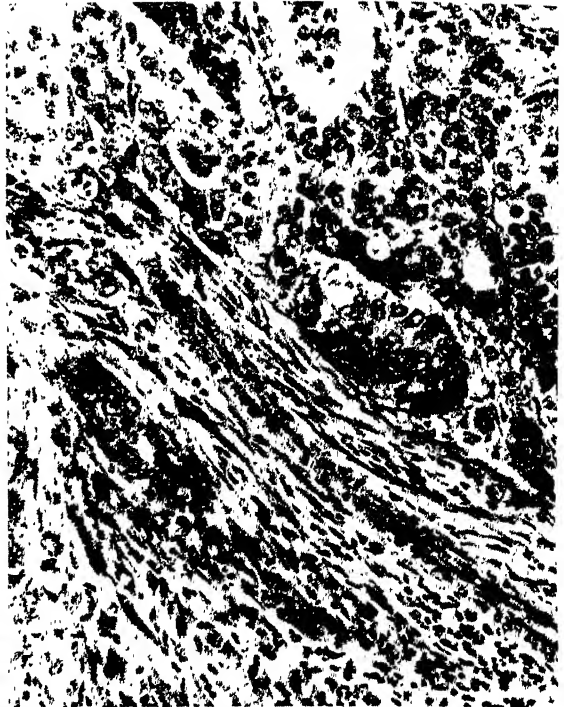


Fig. 12.15 Adenocarcinoma of bowel invading the muscle coat. *Below*, nearly solid strands of tumour are invading smooth muscle on each side of the arteriole which runs up from the bottom right corner. *Above*, the older tumour strands are developing gland-like lumens. $\times 160$.

tion, and perhaps partly because the gland is usually in a non-secretory state.

(b) **Cystadenocarcinoma**, in which cysts lined by columnar or cuboidal cells are prominent. This is common in the ovary and is seen occasionally in the pancreas and kidney.

(c) **Papillary adenocarcinoma**, in which papillary processes project into cysts. This is seen particularly in the thyroid, ovary and biliary tract.

(d) **Mucous (or mucoid) carcinoma**, an adenocarcinoma in which mucus secretion is unusually marked. The term should be used only when the whole tumour looks like a mass of jelly, and under the microscope most of the tumour cells float free in lakes of mucus (Fig. 12.17). The commonest site of origin in this country is the colon, but it occurs also fairly often in the stomach and breast.

Hard and soft carcinomas

Classification of carcinomas into the two following types depends on features of the stroma

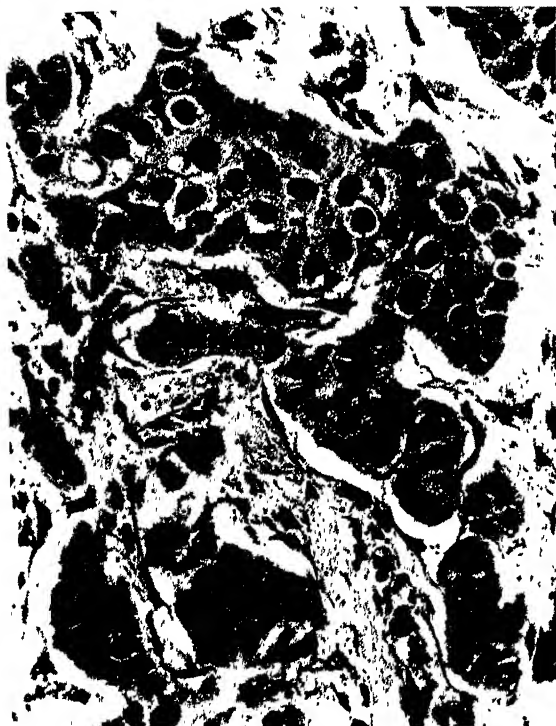


Fig. 12.16 Spheroidal-cell carcinoma of breast infiltrating tissue spaces. The tumour cells are still obviously epithelial, and nuclear changes are not gross, but there is no trace of glandular differentiation. $\times 300$.

and not of the tumour cells. It is thus an essentially different mode of classification. For example both spheroidal-cell carcinoma and adenocarcinoma can be either scirrhus (hard) or encephaloid (soft).

(1) **Scirrhus carcinoma*** is one that shows a very dense fibrous reaction (Fig. 12.18) which is responsible for its hardness. The term is used most often for breast cancer and for the curious 'signet-ring' cell carcinomas of the stomach (Fig. 19.41, p. 617).

(2) **Encephaloid carcinoma** (Fig. 12.19) has minimal stroma and so is soft and 'brain-like' to the touch. The term is rarely used except for the uncommon soft carcinomas of the breast.

These two types appear and feel very different, but behave in much the same way.

Special types of carcinoma

The names of some carcinomas reflect a striking appearance linked to a distinctive behaviour. Examples include clear-cell carcinoma of the

**Scirrhus*: a hard swelling.



Fig. 12.17 Mucous carcinoma. The tumour cells, still in this case with some traces of glandular arrangement, lie in large pools of mucin. $\times 65$.



Fig. 12.18 'Scirrhus' carcinoma of breast, with dense poorly cellular collagen between the tumour cell groups. In this case, in contrast to Fig. 12.16, some of the cell groups (bottom right) show some glandular differentiation. $\times 240$.

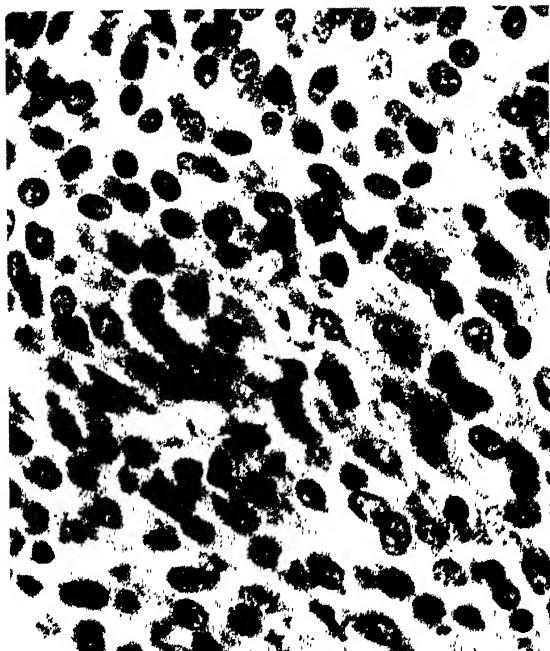


Fig. 12.19 'Encephaloid' carcinoma of breast, showing cells essentially similar to those of Figs. 12.16 and 12.18, but in large masses with very little stroma. Only part of a large mass of cells is shown here. $\times 525$.



Fig. 12.20 Carcinoma arising from liver cells, showing trabecular arrangement. $\times 190$.

kidney, hepatocellular carcinoma (Fig. 12.20), choriocarcinoma of the placenta, and rodent ulcer of the skin; these are highly distinctive lesions with very marked peculiarities of histogenesis as well as appearance and behaviour. They are described in the appropriate systematic chapters.

Spread of carcinoma

Local invasion

Local invasion by a carcinoma depends in part on the sheer expansive pressure of the mass of growing cells, but also on the active migration of motile tumour cells which penetrate the surrounding tissues and then multiply at the new site. Sometimes they appear to migrate as single cells: more often they appear to penetrate as columns of cells which extend by growth at the forward end. We know little about why cancer cells behave thus, but, as indicated in the last chapter, it may be due to increased motility, or to loss of cell adhesiveness or of other normal restraining processes.

Growth is easiest along the planes of loose connective tissue, and may be checked by dense

structures such as thick fascia, the walls of large arteries, cartilage and compact bone. Structures such as glands and muscle, once penetrated, are rapidly destroyed, partly by pressure, partly by loss of blood supply. The carcinoma itself tends to outrun its blood supply, and necrosis of the tumour which ensues includes ischaemic destruction of any normal tissues which have been invaded. Unfortunately the growing edge of the tumour is hardly ever included in the necrosis.

Local invasion, may, of course result in damage to major structures nearby. Of greater significance from the point of view of the life of the patient in most cases is the appearance of **secondary deposits** or **metastases**—new areas of growth of the tumour at a distance from the primary tumour. These result from spread of tumour cells from the primary growth, usually by the lymphatics or blood vessels, as described below.

Spread by lymphatics

This is one of the most characteristic features of carcinoma and is of prime importance from

the surgical point of view. Cancer cells penetrating into the lymphatics may either float free in the lymph and be arrested in the lymph nodes, or they may (probably less often) form columns of proliferating cells filling and growing along the lymphatics (Figs. 12.21, 12.22). Small nodules of tumour may be formed along the line of the lymphatics, but the largest nodules are those which form in and replace the lymph nodes. Early node metastases usually lie in the peripheral sinus (Fig. 12.23). The lymph nodes draining the region of the carcinoma are usually first and most extensively involved. Occasionally spread may take place in a direction contrary to normal lymph flow (Fig. 12.24) as a sequel to lymphatic obstruction.

By excising a carcinoma together with the surrounding tissues, it is often possible to remove completely the local growth. But there may be extensive permeation of lymphatics by cancer cells, although often not visible to the naked eye (Fig. 12.25). Accordingly tissues



Fig. 12.21 Lymph spread of carcinoma, involving both lymphatics and lymph nodes, around the bifurcation of the aorta. Lymphatics filled with tumour are particularly well seen as they cross the left common iliac artery just below right centre of the specimen. $\times 0.75$.



Fig. 12.22 Lymphatic spread of carcinoma. This section of lung shows gross distension of perivascular lymphatics (normally scarcely seen at this magnification) by cancer. $\times 12$. The primary tumour was a carcinoma of stomach, seen at higher magnification (inset) to be composed of mucus-secreting 'signet-ring' cells (see p. 617).

which appear grossly normal may be extensively involved.

The presence of these minute foci of cancer cells, extending beyond the main mass of the tumour either by direct invasion or via the lymphatics, explains the local recurrence of cancer after surgical removal. **Recurrence** is nearly always the result of growth of cancer cells which have been left behind in the surrounding tissues. Such cells may remain dormant, so that years or even decades may elapse before a recognisable tumour reappears. Because carcinogens often affect a large area of epithelium, e.g. chemicals affecting the urinary tract epithelium (p. 299) or ultraviolet rays affecting the exposed skin (p. 301), successful removal of a carcinoma may be followed by the development of a second one nearby. It is usually not possible to distinguish this from recurrence, although the detection also of premalignant change (p. 337) in the epithelium concerned makes it more

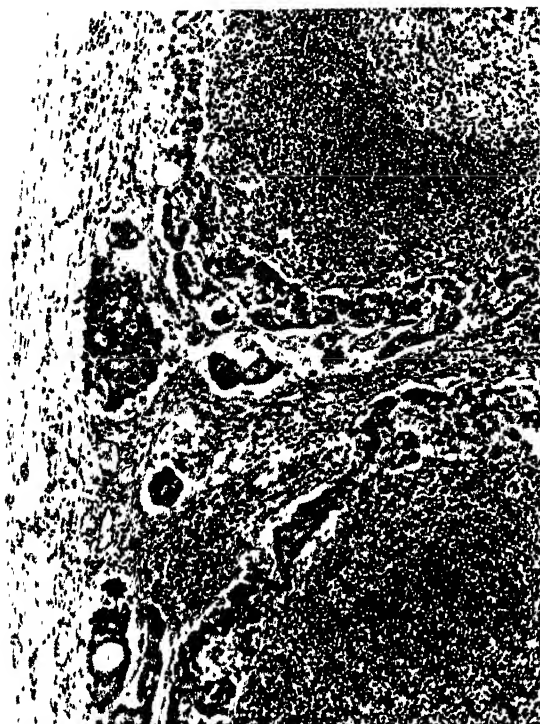


Fig. 12.23 Carcinoma invading lymph node. Carcinoma cells are seen in the lymph vessels in the capsule of the node and in the peripheral lymph sinus, from which they are extending into the radial sinuses. $\times 75$.

likely that the second tumour is not a recurrence of the first.

Blood spread

With most carcinomas, the effects of blood spread are seen later than those of lymphatic spread, though the wider dissemination makes them usually of more serious immediate consequence to the patient. The small veins in and around the primary tumour are the usual route of entry to the circulation. Tumour cells are then carried away to lodge in the next capillary bed that the blood passes through, e.g. in the liver if the primary tumour is in the portal drainage area (Fig. 12.26), and in the lungs from tumours in most other sites. Thence they may spread further, from the liver to the lungs and from the lungs via the systemic circulation to any part of the body (Fig. 12.27). Blood-spread metastases in a solid organ present a very characteristic picture of multiple rounded nodules, scattered randomly throughout the organ and varying in size but usually with no single nodule conspicuously larger than the rest (Figs. 12.28 and 20.55, p. 703).



Fig. 12.24 Retrograde invasion of lymph node by carcinoma. The lymphatics at the hilum of the node are filled with cancer cells, which have spread into the node against the normal direction of lymph flow. $\times 95$.



Fig. 12.25 Squamous carcinoma of vulva, showing lymphatic permeation in the dermis beyond the clinically apparent margin. The larger tumour masses (above right) would appear to the naked eye to be the edge of the tumour, the bulk of which lies further to the right. Lymphatic permeation can usually only be recognised with certainty in microscopic sections, though when it is as gross as that seen here it may be suspected on careful examination with the naked eye. $\times 15$.



Fig. 12.26 Secondary carcinoma in the liver. Note masses of cancer cells in the sinusoids between the (darker) liver cells, without any formation of stroma. \times about 200.



Fig. 12.27 An embolus of carcinoma cells in a glomerulus, with extension into the tubule. \times 150.



Fig. 12.28 Secondary carcinoma in bone. Multiple rounded white masses in the humerus, in a case of carcinoma of breast. \times 0.75.

There are in practice many apparent anomalies in the distribution of metastases: some may be due to confusion between blood and lymph spread—there is for instance a strong case for regarding the curious predilection of bronchial carcinoma to spread to the adrenals as a consequence of lymph spread rather than (as has been generally believed) blood spread. But even allowing for this, there must be great variations between capacities of different tissues to resist the growth of tumour cells arriving by the bloodstream. It is highly probable that in most cases the great majority of cells leaving the primary tumour by the veins fail to establish themselves.

In his old but still valuable series of necropsies on patients with carcinoma, Willis (1973) gave the following incidences of metastases in various organs:

liver	36%	brain	6%
lungs	29%	spleen	3%
bones	14%	skeletal muscles	1%
adrenals	9%	skin	1%

These incidences clearly do not correspond with blood flow, and so with the number of tumour cells likely to be arriving in the various organs. The liver and bone marrow are susceptible sites: the spleen and muscles resistant. There is experimental evidence for destruction of tumour cells by the spleen, but the nature of the defence mechanism is not known. The lung, which must receive by far the largest number of tumour cell emboli, provides an environment which is only moderately favourable to their growth. Cases of carcinoma in some organs, for instance kidney, prostate and thyroid, often have multiple systemic metastases, for example in bone, with no obvious lung lesions: in some such cases it can be shown that there are minute microscopic foci in the lung where tumour cells have lodged in the pulmonary vessels, and, while failing to grow to any substantial extent at the site, have been able to launch further tumour emboli into the systemic circulation (Fig. 12.29).

Retrograde venous spread. One type of anomaly that has a special explanation is the localisation of metastases within the axial skeleton. Carcinoma of the prostate spreads early

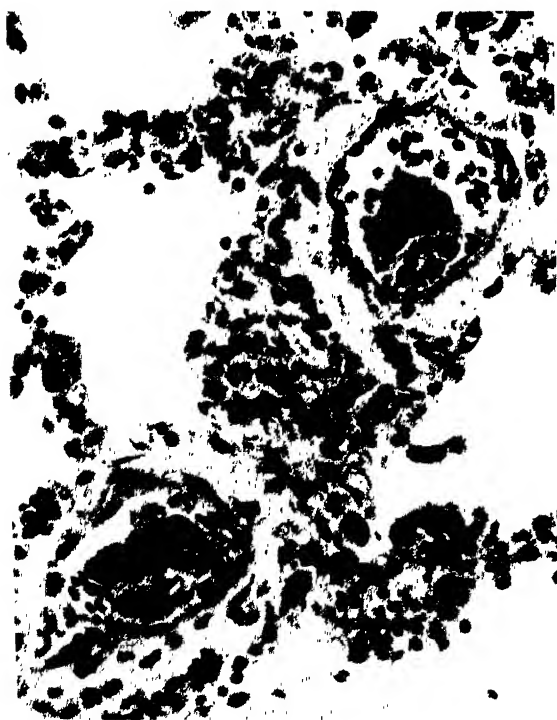


Fig. 12.29 Lung from a case of prostatic carcinoma, showing two pulmonary arterioles containing tumour cells. $\times 190$.

to the lumbar spine and pelvis, carcinoma of the breast to the thoracic vertebral bodies, and carcinoma of the nasopharynx to the cervical spine and the base of the skull. This is not attributable to direct or lymph spread but to spread by the blood. Yet blood spread in the usual fashion via the lungs should result in uniform spread to all parts of the vertebral column. The explanation for the localised involvements of the spine appears to lie in the peculiarities of blood flow in the intra-vertebral venous plexus, in which differences of pressure above and below the diaphragm often lead to reversal of flow: the effect is to draw venous blood at times into the vertebra from neighbouring organs, and this may carry malignant cells, which find a very favourable site for multiplication within the vertebral marrow.

Intracavitary spread

When carcinoma involves a body cavity, cancer cells may be liberated into the space and graft themselves on the surface to form new foci of growth. Any cavity can be involved, and the *subarachnoid space*, for example, is of some importance in the spread of intracranial tumours, but the serous cavities are especially important for carcinoma. Involvement of the *peritoneum* results most often from carcinomas arising in the stomach or the ovary, while the *pleura* and *pericardium* are most commonly invaded by carcinomas of breast or bronchus. In most cases there is an effusion of fluid into the sac concerned, and this may be bloodstained. Malignant cells are often present in such an effusion, but very often it is difficult to distinguish them with certainty from altered serosal cells. A special example of this form of spread is seen in transperitoneal metastasis to the ovary, usually before the menopause and usually from a gastric carcinoma of the 'signet-ring' type (p. 617): the ovaries may become very large, and have the characters first described by Krukenberg, who thought that such tumours originated in the ovaries (see p. 966).

Although not everyone agrees, it is nearly certain that the rare so-called 'alveolar-cell carcinoma' of the lung (p. 498) grows by spread of carcinoma cells within the air passages of the lung.

Intra-epithelial and intracellular spread

Cancer cells may invade the epidermis, which is the only extensive epithelium in the body thick enough to withstand such invasion without disruption. Cells may spread for several centimetres in the epidermis without destroying it completely. Paget's disease of the breast (p. 983) is the only common example involving carcinoma cells (the tumour being derived from the underlying breast): rarely this occurs in the epidermis at other sites, and something similar is seen in malignant melanoma of the skin (p. 1082).

Carcinoma invading skeletal muscle may sometimes be seen under the microscope to be growing within the sarcolemma of muscle cells: this rare curiosity is the only known example of 'intracellular' spread (Fig. 12.30).



Fig. 12.30 Intracellular invasion. The long sausages of tumour cells have grown within and expanded striped muscle fibres, being still confined by the sarcolemma: compare with unaffected muscle fibres on the left. $\times 60$.

Premalignant lesions

These are conveniently considered here, as most such lesions involve epithelium, and are precarcinomatous.

A premalignant condition is one which can be recognised by either clinician or pathologist, and which indicates that the bearer has a substantially greater than normal risk of developing a malignant tumour. The early stages of premalignant change (which presumably indicate the occurrence of the earlier mutations of a multistage conversion to malignancy) can often be recognised histologically. The signs include nuclear irregularity, increased mitotic activity, and abnormalities of differentiation, often combined with inflammatory infiltrates and stromal changes. The risk of such lesions' becoming malignant can be established only in the light of experience of their behaviour in each particular site in which they occur. The following are examples of premalignant lesions.

(a) **Some benign tumours.** Probably all benign tumours carry some increased risk of malignancy, but in most it is very little more than that of the normal tissue, while in others it is high. There is little obvious logic about the differences. The rare polypoid adenomas of the small intestine seldom become malignant, while the (much commoner) adenomas of the large intestine carry a much greater risk of cancer. Comparable anomalies could be quoted at other sites.

(b) **Certain chronic diseases.** Carcinoma may develop as a more or less common complication of some non-neoplastic diseases, such as cirrhosis of the liver, ulcerative colitis, asbestosis and a variety of skin diseases.

(c) **Carcinoma-in-situ.** In the most extreme degrees of premalignant change in an epithelium, all the cytological changes of malignancy are seen in its cells, but the altered cells remain in the intact layer of epithelium and do not invade the underlying tissues. The name carcinoma-in-situ is often given to this lesion: it is not strictly accurate, as no carcinoma is present until invasion begins, but it is a vivid reminder of the need for action. *Intraduct carcinoma* of the breast (p. 982) is an essentially similar lesion.

Carcinoma-in-situ of the cervix uteri is particularly important, for the cervix is accessible for examination, the lesion is relatively

common, and is readily detected by cytological examination of smears and confirmed by biopsy. Also, in most cases the in-situ change persists for years without becoming invasive, allowing its detection by routine screening and prevention of cancer by excision of the affected epithelium (which varies widely in extent).

Although frank invasive cancer in many other sites is preceded by carcinoma-in-situ, detection at this stage is more difficult because of inaccessibility, etc.

The degree of risk of developing malignancy varies greatly with different premalignant lesions, and is often hard to determine with precision. In a few conditions, such as polyposis (adenomatosis) coli and xeroderma pigmentosa, it is practically 100 per cent. In carcinoma-in-situ of the cervix, the rate remains uncertain despite extensive studies, but it is believed that about 30 per cent progress to invasive carcinoma. In most other lesions the risk appears to be much lower.

Staging and grading of cancers

A quantitative measure of the factors affecting the prognosis of a particular type of tumour is often required. Sometimes it is used in deciding on the best method of treatment, but it is most useful in statistical studies. If, for instance, a surgeon claims that his new operation is curing more patients than his colleagues (or, to be more precise, raising the proportion of five-year survivals) it is necessary to be sure that he is not seeing by some accident an exceptionally favourable group of cases—e.g. smaller, earlier or better differentiated tumours. The production of completely unambiguous criteria is much more difficult than it sounds, and elaborate special systems have been developed for most of the common tumours. All agree, however, in separating two main elements, *grading* and *staging*, and these should not be confused.

Grading. This is a *histological* estimate of the degree of *differentiation*. Exact numerical systems do not work (Broder's system, depending on the percentage of differentiated cells, though still often talked of, has long been abandoned

in practice). In general, most tumours are graded for this purpose by the pathologist into well differentiated, average, and poorly differentiated: the system is most useful if the criteria used put about 25 per cent of the tumours into the good group, 50 per cent into the average and 25 per cent into the worst.

Staging. This is a *clinical* estimate of the degree of *spread*. Most systems use four stages, roughly definable as (I) confined to the organ of origin, (II) local spread not interfering with surgical excision, (III) fixation to surrounding structures and (IV) distant spread. A 'Stage O' is sometimes added for invasive tumours of microscopic size or for carcinoma-in-situ: the difference between the two is, however, important and, if used at all, 'Stage O' should be clearly defined.

How this works in practice may be seen from the usual definitions for cancer of the uterine cervix: *Stage O*—carcinoma-in-situ only: *Stage I*—confined strictly to the cervix: *Stage II*—local spread, not reaching the pelvic wall or the lower third of the vagina: *Stage III*—fixed to pelvis, or involving lower third of vagina: *Stage IV*—distant metastases, or involvement of bladder or rectum. Stage II is often subdivided, IIa including spread to the uterine body, vaginal fornices and the immediately adjacent connective tissue, Stage IIb including further spread, but short of Stage III. It will be seen that even here the position is not altogether simple. For most other tumours staging is a good deal more complicated and often controversial.

Staging must always be based primarily on the clinical examination, for the assessment at operation often differs from the clinical assessment, and to compare surgical with other forms of treatment, both groups of patients must first be 'staged' in the same way, i.e. clinically. Grading, on the other hand, requires pathological examination of at least a biopsy. The results of the two procedures are not, however, altogether independent: as one might expect, in all series the worse differentiated cases tend to be more numerous in the higher stages.

Further Reading

- Ashley, D. B. (1978). *Evans' Histological Appearances of Tumours*, 3rd edn., pp. 857. Churchill-Livingstone, Edinburgh, London and New York. (An account of the behaviour and appearances of human tumours based on a considerable experience.)
- Sobin, L. H., Thomas, L. B., Percy, Constance and Henson, D. E. (Eds.) (1978). *A Coded Compendium of the International Histological Classification of Tumours*, pp. 116. World Health Organisation, Geneva. (A widely accepted system of classification, coding and nomenclature of human tumours.)
- Willis, R. A. (1973). *The Spread of Tumours in the Human Body*, 3rd edn., pp. 417. Butterworths, London.
- Atlas of Tumour Pathology*. US Armed Forces Institute of Pathology, Washington, DC. (Numerous 'Fascicles' on tumours of particular organs, tissues and regions. A valuable source of detailed information on the histology and behaviour of individual tumours.)
- Cancer*. A journal of the American Cancer Society. Lipincott, Philadelphia and Toronto. (A monthly publication of well-illustrated articles on human neoplasms.)

Tumours: III. Other Varieties

There are far more kinds of non-epithelial tissue than epithelial, and equally there are far more kinds of non-epithelial tumour. As already indicated, however, the balance between this chapter and the last reflects the practical circumstance that the carcinomas greatly outweigh all other tumours in clinical importance. Nevertheless, even the tenth of deaths that are due to non-epithelial malignant tumours is a substantial number, and no one can dismiss as unimportant a group that includes the lymphomas, gliomas, mela-

nomas and bone sarcomas. There are moreover some benign tumours, e.g. the myomas, and the tumour-like angiomas, which are very common, although responsible for few deaths.

Though all the main types of non-epithelial tumours will be found mentioned here, detailed description will be given only of those varieties which are of such general distribution as not to be easily included under any one system. Length of description does not therefore always reflect importance.

Tumours of the Connective Tissues

Nomenclature

In this group it is usual to name tumours by adding *-oma* to the appropriate stem for the benign lesion, and *-sarcoma* for the malignant. Thus *fibroma* and *fibrosarcoma* are respectively benign and malignant tumours arising from fibrocytes; *chondroma* and *chondrosarcoma* are benign and malignant tumours arising from cartilage cells.

Benign tumours

These are composed chiefly of fully developed tissues, such as are found in the adult body, e.g. fibrous tissue, cartilage, muscle. They are usually rounded or lobulated and well defined, being generally enclosed within a distinct fibrous capsule. They displace the surrounding tissues and produce atrophy by pressure, but they do not usually infiltrate tissues and never metastasise. Such benign tumours are sometimes multiple, but then each tumour represents an independent focus of growth. Blood vessels grow in relation to the tumour tissue, and are

usually well formed, though the arteries are often deficient in muscle fibres. The well-defined reactive fibrous tissue stroma of epithelial tumours is, however, altogether absent in most cases, the tumour relying for its support on the matrix produced by its own cells—though exceptions occur, for example in the myomas.

Malignant tumours (sarcomas)

In the corresponding malignant tumours, the activity of the cells is mainly proliferative, the tumour is more cellular, and although a certain amount of matrix is formed, this is usually scanty. Sarcomas often form large masses, usually soft and commonly with areas of haemorrhage and necrosis (Fig. 13.1). While these malignant tumours may appear to the surgeon to be encapsulated, as in Fig. 13.1, diffuse destructive infiltration of the surrounding tissue occurs at the margin of most such masses so that wide excision rather than enucleation is required in their treatment. There is an exten-

Tumours of fibrous tissue

Although it is usual to regard these as the 'typical' connective tissue tumours, they are in fact not very common and genuine benign fibromas are rare.

Fibroma

This is the name given to benign tumours of fibrous tissue. The cells of the tumour are fibrocytes; their nuclei are long and narrow and densely staining, the cytoplasm so scanty as to be hard to see, and mitoses are very rare. Bundles of dense collagen separate the cells, and the appearances, in short, may not be very different from that of fibrous tissue as seen in a thick fascia or a scar. Fibromas do, however, vary greatly in the amount of collagen and in cellularity (Figs. 13.2, 13.3).



Fig. 13.1 Spindle-cell sarcoma arising in intermuscular fascia of thigh. Above and left there is normal voluntary muscle. The upper half of the ovoid tumour shows extensive necrosis and haemorrhage: the lower half is well preserved, appearing characteristically greyish-white, soft and slightly lobulated. $\times 0.6$.

sive new formation of poorly formed blood vessels. Numerous capillaries, and larger channels composed mainly of a layer of endothelium, are supported by the cells of the tumour, while fibrous tissue is found only round the larger vessels. Two results follow—(a) the cells of the tumour readily break through the vessel walls and are conveyed in the venous blood until arrested in the smaller vessels of lungs, liver, etc., and thus **metastases** may develop, and (b) **haemorrhages** are common. Spread by the lymphatic vessels is unusual, except in the case of synovial sarcoma and lymphoma, and blood-borne metastases, especially in the lungs, are the usual cause of death.

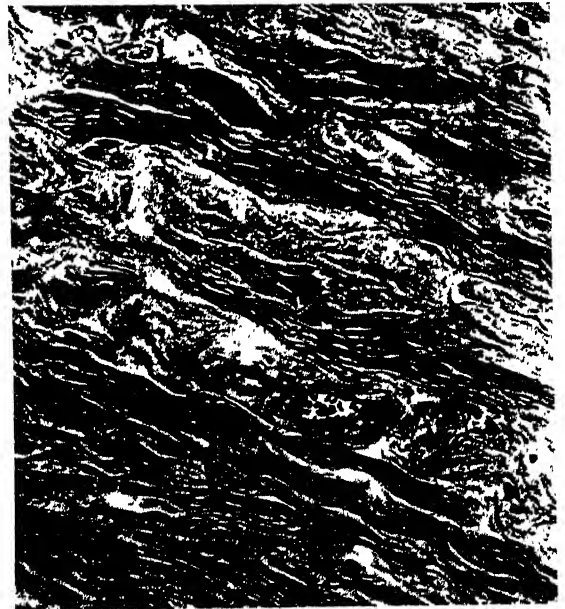


Fig. 13.2 Fibroma showing abundant mature collagen and few cells. Distinction from normal fibrous tissue can be difficult, and depends on careful examination of the whole lesion. $\times 240$.

Fibromas may occur in any type of connective tissue (Fig. 13.4) and may be seen with varying degrees of rarity in most of the internal organs. Special varieties occur in the sheaths of nerves (Schwannoma, etc., pp. 793–4), skin (dermatofibroma, p. 1084) and ovaries (p. 965) which have their own peculiarities of behaviour

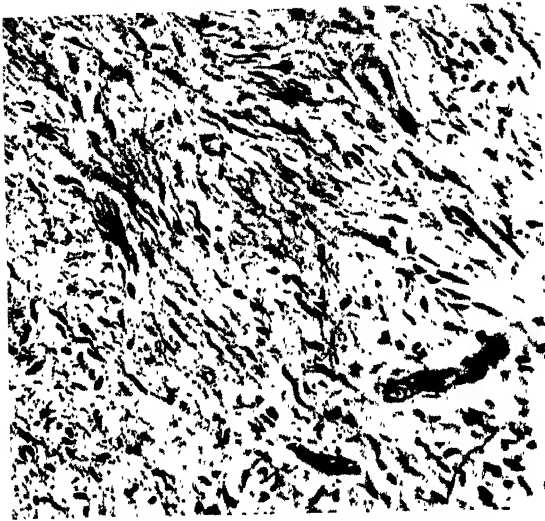


Fig. 13.3 Fibroma showing more numerous cells (which are however mature fibrocytes) and less abundant and looser collagen. $\times 280$.



Fig. 13.4 Lobulated fibroma removed from buttock. Natural size.

and which probably do not arise from ordinary fibrocytes.

Fibromatoses

This term can be usefully applied to some fibroma-like lesions which may cause considerable difficulties of diagnosis and which may not be true tumours. They include:

(a) **Desmoid tumour**, which is a curious lesion seen characteristically in the rectus abdominis muscle of multiparous women. It has the histology of a fibroma, but is not encapsulated and infiltrates the surrounding muscle and destroys the muscle fibres (Fig. 13.5). It often recurs locally after excision, but never metastasises. Desmoids occur also in the thigh and shoulder, where they may be termed **musculo-aponeurotic fibromatosis** and behave similarly. Genuine fibromas are particularly rare in skeletal muscle.

(b) **Palmar fibromatosis (Dupuytren's contracture)** consists of a fibroma-like lesion, often quite cellular, involving the palmar fascia and producing flexion deformities of the fingers (see p. 929). A similar but much less common lesion of the plantar fascia (*plantar fibromatosis*) is often particularly cellular and sarcoma-like under the microscope but does not cause contracture. There is also a penile fibromatosis (*Peyronie's disease*). Two or even all three of these conditions may occur in one patient, and they are clearly related. Again, none of these ever metastasises.

(c) **Keloid**. Some people (negroes more often than others) have a curious tendency to pro-

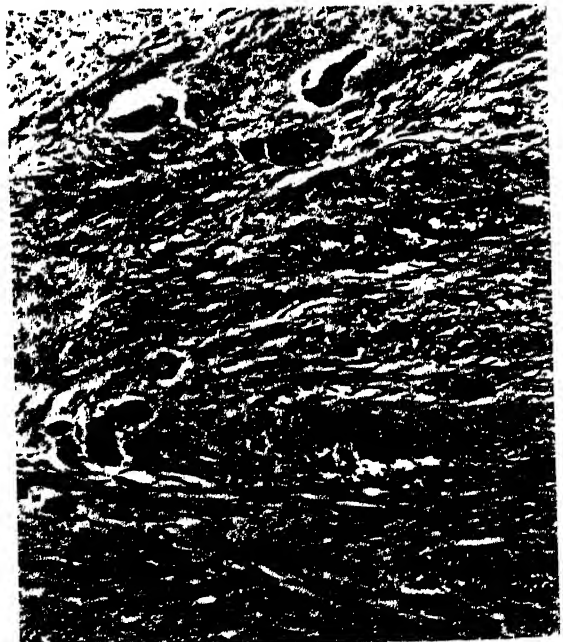


Fig. 13.5 'Desmoid tumour' of rectus sheath. Moderately cellular fibrous tissue forms the bulk of it. The large dark-staining cells are multinucleate sarcolemmal giant cells, the remains of infiltrated muscle fibres. $\times 120$.

duce excessive amounts of dense hyalinised fibrous tissue, instead of the normal inconspicuous scars, after injury to the skin. They nearly always cease growth after a time, and are then clearly not tumours, though they may be mistaken for them histologically.

Myxoma and mesenchymoma

The rare myxomas (Fig. 13.6) are translucent tumours, usually benign, composed of 'myxoid' tissue, a kind of connective tissue with stellate cells widely separated by a ground substance rich in mucopolysaccharide (p. 247) and poor in blood vessels. They usually occur within skeletal muscles and only rarely recur; even after incomplete excision. There is a group in which a predominantly myxomatous tumour contains scattered elements of other mesenchymal tissue types, including muscle and cartilage. The name **mesenchymoma** is usually given to them, reflecting a belief, not necessarily well founded, that they represent a return to the capacity for multipotent differentiation of primitive mesenchyme. These tumours, most often seen in the subcutaneous tissues of the trunk, rarely metastasise but have a high incidence of local recurrence. The *embryonal rhabdomyosarcoma* of children (p. 346) also combines myxoid connective tissue with poorly formed striated

muscle cells but has a very different age and anatomical distribution and is much more malignant.

Fibrosarcomas

These tumours arise especially from fascia and deep connective tissues, but may occur almost anywhere in the body. Similar tumours arise from nerves—**neurofibrosarcomas** or **malignant schwannomas** (Fig. 13.7). While most fibrosarcomas are clearly malignant from the start, some progress over a period of many years from an early stage in which they may be difficult to distinguish from a benign fibroma.

Fibrosarcomas differ greatly in their degree of differentiation. *Low grade fibrosarcomas* differ little from cellular fibromas: they are firm and fibrous, produce abundant collagen, and the cells differ little from normal fibroblasts. A moderately high rate of mitosis is often the only real evidence of malignancy. Such tumours are slow growing and are often cured by adequate local excision, and recurrences tend, at least at first, to remain localised. Tumours of intermediate malignancy are often called 'spindle-cell sarcomas' (Fig. 13.8). They are softer, more rapidly growing tumours in which the cells are still recognisably fibroblast-like and regularly arranged, but collagen is relatively inconspicuous: such tumours often metastasise and are usually ultimately fatal.

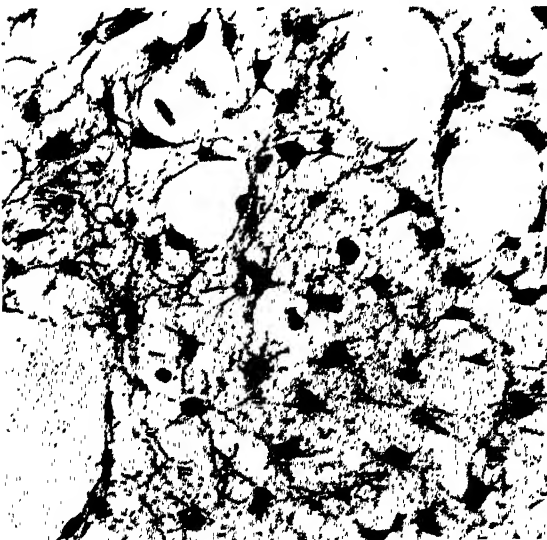


Fig. 13.6 Myxoma. The branching cells lie in a semi-fluid matrix which stains pale lilac with haematoxylin and eosin but much more deeply with stains for connective tissue mucin. The round spaces are included fat cells of the breast, in which this tumour was found. $\times 300$.



Fig. 13.7 Well-differentiated neurofibrosarcoma arising in recurrent neurofibroma. Note the very pronounced palisading of the nuclei. $\times 270$.

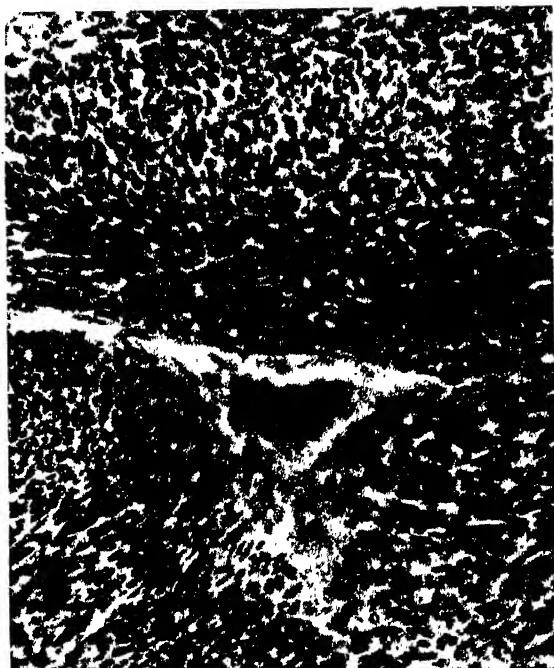


Fig. 13.8 Spindle-cell sarcoma. This is very cellular, but the cells are relatively uniform and still resemble fibroblasts. Special stains would show sparse collagen fibres. The triangular cleft is one of the poorly formed blood vessels characteristic of sarcomas. $\times 250$.

From this type, transitions occur to the anaplastic sarcoma (Fig. 13.9), a soft, rapidly growing tumour consisting of large irregular cells with large irregular nuclei, with little or no collagen and very little evidence of the tissue of origin; metastasis is usually rapid and prognosis poor. Obviously, in such tumours it may be impossible to identify the cell of origin, and indeed tumours with this kind of histology may arise from dedifferentiation of almost any kind of cell.

Fibrosarcomas, like most other sarcomas, may recur locally or metastasise by the bloodstream, especially to the lungs, but rarely spread by the lymphatics.

Tumours of adipose tissue

Lipoma

This is a common, benign tumour which consists of adipose tissue. It increases in size by proliferation of fibroblast-like cells which lie around the blood vessels but are hard to see in most cases because they rapidly accumulate fat

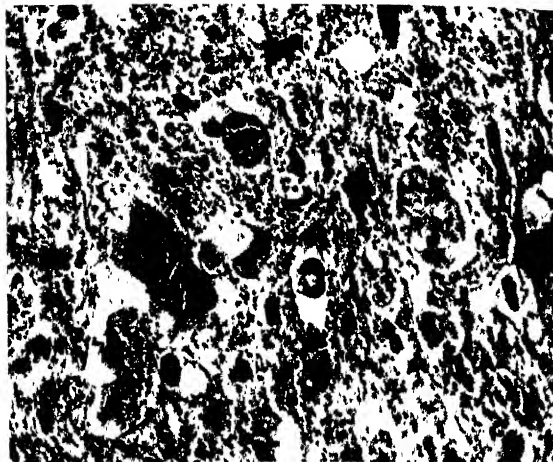


Fig. 13.9 Anaplastic sarcoma showing great variation in the size of the cells. Note the numerous enormous polyploid nuclei. $\times 125$.

and become lipocytes. The common lipoma is a rounded, well-demarcated, subcutaneous mass. It sometimes reaches a considerable size, and may have blunt projections which pass into the tissues around. Multiple tumours may be present and occasionally they are symmetrical. They are commonest over the neck, back and shoulders and occur also in the retroperitoneum, especially in the perirenal fat. They may, however, arise almost anywhere in the body. Should the patient become emaciated, the fat in the tumour is not utilised—a good example of the failure of tumours to respond to the factors controlling the metabolism of normal tissues. In a lipoma there may be areas of fibrous or capillary angiomatous tissue—the tumour being then called a *fibrolipoma* or *angiolipoma* respectively; occasionally there are areas of calcification.

In a rare variant the fat is finely divided in droplets within the cells, which thus closely resemble brown fat. The supposed role of brown fat in hibernation of rodents led to the name '*hibernoma*' (Fig. 13.10).

Liposarcoma

This tumour is uncommon, although one of the least rare of the soft tissue sarcomas. It occurs most often in the thigh, buttocks and retroperitoneum. It varies widely in naked-eye and microscopical appearances and also in prognosis.

The best differentiated liposarcomas are

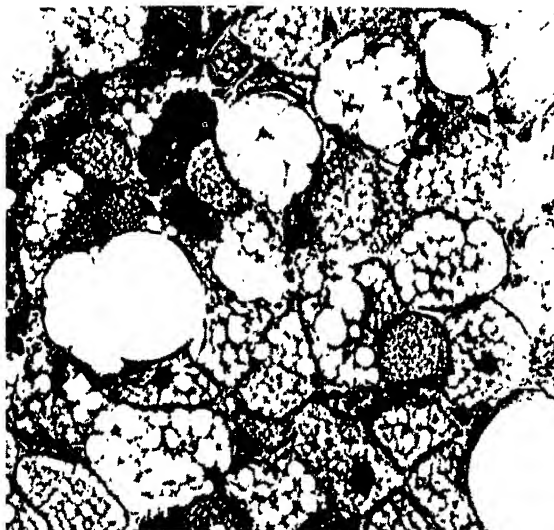


Fig. 13.10 'Hibernoma' showing the characteristic appearances of the fat-laden cells, with central nuclei. $\times 480$.

obviously fatty, and have a marked microscopic resemblance to lipoma, but with more obvious collagenous areas and groups of cells with larger and more hyperchromatic nuclei.

The least uncommon type, the *myxoid liposarcoma*, appears gelatinous, having a mucopolysaccharide-rich matrix, with a prominent capillary network (Fig. 13.11). It has a marked

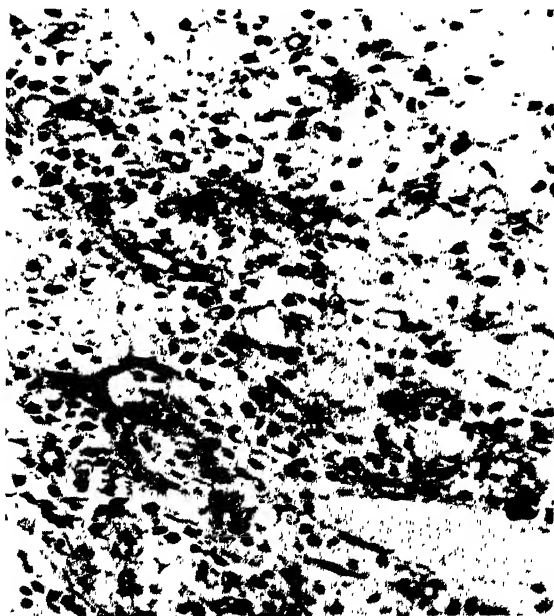


Fig. 13.11 Myxoid liposarcoma, showing the relatively small sarcoma cells, with scanty cytoplasm, lying in an abundant homogeneous matrix. Note the numerous capillaries. $\times 250$.

tendency to repeated recurrence after successive attempts at removal, and sometimes recurs in a more malignant form, consisting of round cells containing little fat and with much less stroma. The round-cell liposarcoma and another type of liposarcoma consisting of cells containing abundant fatty droplets but showing extreme cellular pleomorphism (Fig. 13.12) are liable to metastasise.

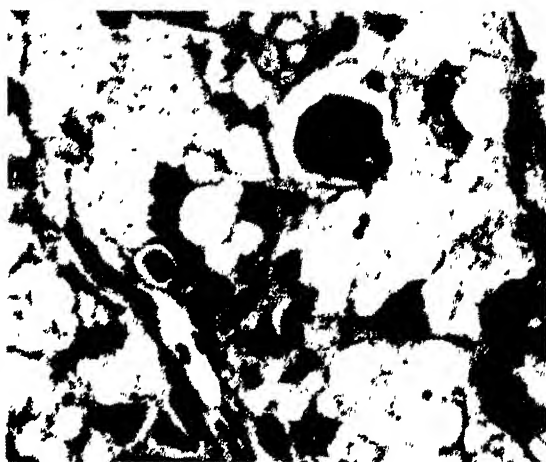


Fig. 13.12 Pleomorphic liposarcoma, consisting of very large cells with very large nuclei, and with much fat in the cytoplasm. $\times 450$.

Alveolar soft tissue sarcoma. This rather uncommon variety of sarcoma arises in the soft tissues, usually of a limb. It is composed of large polygonal or round cells with coarsely granular cytoplasm, and arranged in a curiously alveolar pattern. In paraffin sections the appearances resemble those of liposarcoma or even carcinoma but the granules do not give the staining reactions of lipids, mucin, glycogen or other specifically stainable substances. Its origin and true nature are obscure, but some regard it as a variety of chemodectoma (p. 349).

Tumours of cartilage and bone

These are dealt with in detail in Chapter 23. The benign tumours of bone are a perplexingly varied group, and their precise histogenesis is often obscure. Most of the masses of cartilage called chondromas and many bony outgrowths (exostoses) are not true tumours but developmental defects or hamartomas (p. 358). **Osteosarcomas** particularly and **chondrosarcomas** are among the commonest sarcomas. Osteosarcoma especially exemplifies many of the most characteristic features of sarcomas generally—a high incidence in childhood, high mortality, rarity of

lymphatic spread and frequent blood spread, especially to the lungs.

Fibrosarcomas may arise from bone. Conversely (though very rarely) bone-forming osteosarcomas sometimes arise from connective tissue elsewhere, and sometimes even from such organs as the breast, kidney, etc.

Bone and cartilage are sometimes seen in non-bony tumours. Both are common in teratomas (p. 355). Cartilage is often seen in the mixed tumours of the parotid and other salivary glands. Bony metaplasia of the fibrous stroma of carcinomas is a rare but striking finding, least rare in man in large-bowel cancers, and not uncommon in breast tumours in bitches.

Tumours of muscle

There are two varieties of myoma, the **leiomyoma**, composed of smooth muscle fibres, and the **rhabdomyoma** of striped muscle. The latter is so rare that the term **myoma** without qualification is often used to signify leiomyoma.

Leiomyomas are composed of smooth muscle cells orientated in a more or less parallel manner within bundles, which are arranged in a whorled pattern (Fig. 13.13). A small amount of supporting fibrous tissue runs among the individual cells, while broader bands separate the bundles. The proportion of fibrous tissue to muscle varies much in different specimens. The tumours are usually firm and rounded, and on section are pinkish with a characteristic whorled appearance due to the arrangement of the fibres (Fig. 13.14).

Leiomyomas of the *uterus* are among the commonest of tumours: their usually high content of fibrous tissue earns them their common name of '**fibroids**', though the muscle is the only true tumorous element and it is incorrect to call them fibromyomas. Of general interest is their tendency to cease growth at the menopause and subsequently regress.

Leiomyomas are probably next most common in the muscular coat of the alimentary canal, though here most are too small to be found without special search and few are large enough to cause trouble. They can occur at many other sites, all uncommon: a rare painful vascular myoma of the skin is of special interest (p. 349).

Leiomyosarcomas occasionally arise from the same sites as leiomyomas, especially the uterus

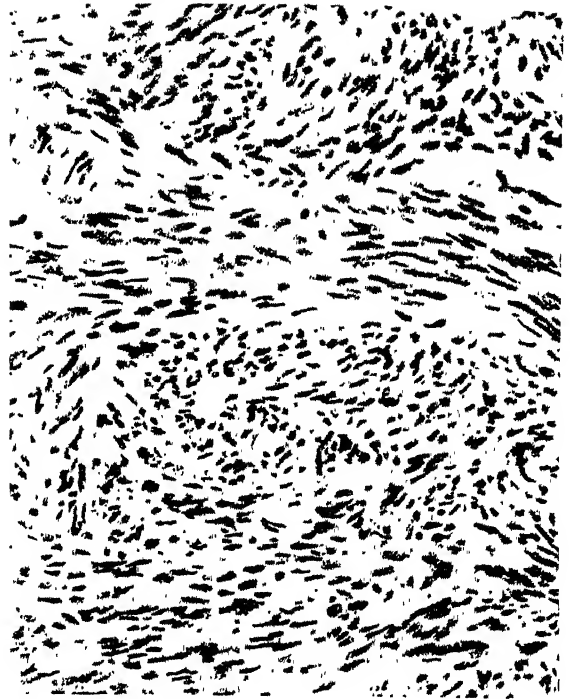


Fig. 13.13 Leiomyoma. The muscle bundles interlace irregularly, and are to be seen cut transversely, longitudinally and obliquely. The cytoplasm of the muscle cells stains somewhat indistinctly but gives a darker shade to the substance of each bundle. Occasional fibrocytes seen between the bundles have smaller darker nuclei and no obvious cytoplasm. $\times 185$.

and the stomach. Most appear to be malignant from the start but some may arise by malignant progression in a benign myoma. The histological diagnosis of malignancy is usually easy on general cytological grounds, but the better differentiated tumours can be hard to distinguish from the more cellular benign tumours. It is a useful empirical rule, which can be applied to few other tumours, that a smooth muscle tumour in which mitoses can be found easily is liable to metastasise.

Rhabdomyosarcoma. In proportion to their bulk, the voluntary muscles are one of the rarest of sites for tumours of any kind, both primary and secondary: the reasons for this are not known. The rare tumours in which striped muscle fibres or their precursors are seen are nearly always malignant, and arise mostly in sites where no striped muscle is present normally. They are found most often in the female genital tract, characteristically in the cervix or vaginal vault in young girls, and in the lower



Fig. 13.14 Leiomyoma of uterus. The tumour is paler than the normal muscle because of its higher content of fibrous tissue. The pattern of the cut tumour surface is formed by brownish slightly translucent strands of muscle, here seen as an indistinct network but often whorled. Note the distorted uterine lumen. Natural size.



Fig. 13.15a Rhabdomyosarcoma. Large pleomorphic cells in a slightly myxoid matrix. Special stains and high magnification show coiled myofibrils (best seen in cells just below and to the right of centre).

urinary tract in both sexes; also in the soft palate. Usually the tumours have a large myxoid element, and recognisable muscle cells (though very characteristic when found) are often scanty and difficult to find, even with special stains which accentuate their cross striations. Rounded or irregular cells with coiled myofibrils but without cross-striations are usually more readily apparent. (Fig. 13.15). When growing beneath a mucous membrane, these tumours often present with numerous blunt translucent processes—hence the name of *sarcoma botryoides* (grapelike sarcoma). These are highly malignant tumours: local recurrence after excision and blood-spread metastases are usual, and (unlike most sarcomas) lymph node metastases are common.

Rhabdomyosarcomas are also seen very rarely in the heart. Benign tumours, perhaps better regarded as hamartomas (p. 358), are seen there in children with epiloia (tuberous sclerosis): they consist of swollen muscle cells packed with glycogen and some confusion has in the past existed between these lesions and the cardiac changes of glycogen storage disease (p. 30).



Fig. 13.15b As 13.15a. The nature of the myofibrils is more obvious when the cells are elongated 'strap cells' and the fibrils produce cross striations as seen here, especially below right, but these are usually hard to find.

Tumours and Malformations of Blood Vessels and Lymphatics

Angiomas

Haemangioma. A haemangioma consists of a mass of blood vessels, atypical or irregular in arrangement and size. A corresponding growth, **lymphangioma**, is composed of lymphatic vessels similarly altered; but, as this is rarer, the term angioma is often used as synonymous with haemangioma.

The majority of the lesions called angiomas are not true tumours, but hamartomas.* They are present at birth, even if not always visible, and their enlargement ceases with the growth of the patient. Most angiomas are well-defined masses of vascular tissue which resemble tumours sufficiently to justify their inclusion here. The two common varieties are as follows.

(a) *Capillary angiomas* consist of dense plexiform arrangements of vessels of capillary size (Fig. 13.16). They occur especially in the skin, where they form one of the two common types of **naevus**† or birthmark, but are also seen in the internal organs. Most are small, but larger lesions occur, e.g. the 'port-wine stains' of the face, which consist of capillary-like vessels with an abnormally large lumen. Capillary angiomas are usually well defined, and deep red or purple. The capillary vessels have a more prominent endothelial lining than normal capillaries and endothelial cells may be seen scattered or in clusters without formation of a lumen (Fig. 13.16). The stroma consists of well-formed collagen. There is no capsule, and outlying groups of capillaries in adjoining tissues often give a false appearance of invasion. The blood supply is usually clearly separated from that of the surrounding tissues, there being generally only one artery of supply.

(b) *Cavernous angiomas* are found in the skin, subcutaneous tissue, lips and tongue and also in the liver. They consist of relatively large interconnecting sinus-like vascular spaces (Fig. 13.17). In the liver they form deep-purple well-defined masses, usually polygonal rather than round and not raised above the surface, signs of their lack of expansile growth.

*A *hamartoma* is a malformation developing in early life and consisting of a tumour-like mass of cells or tissue which grows with the individual and then ceases to grow when general body growth ceases (p. 358).

†A *naevus* (mole or birthmark) is a hamartoma of the skin made conspicuous by some definite colour difference from the surrounding normal area. Angiomas and pigmented naevi (pp. 1078–81) are the two common types..

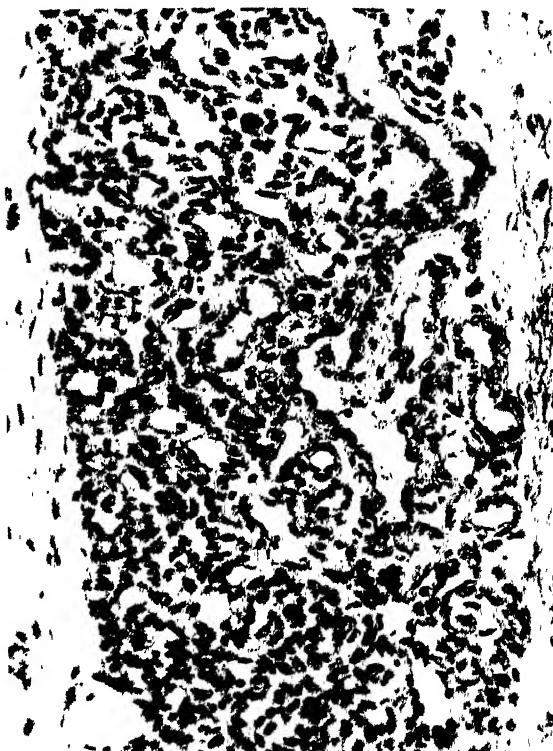


Fig. 13.16 Capillary angioma showing well-formed capillaries with prominent endothelial cells. The solid areas between capillaries include many cells which appear to be endothelial cells not related to a lumen. $\times 200$.

Angiomas are often multiple, and are also an important component of several diseases with a strong genetic predisposition, e.g. hereditary haemorrhagic telangiectasia (multiple small angiomas in skin and mucosae with a strong tendency to haemorrhage—p. 557), Lindau's disease (cerebellar and retinal angiomas with cysts of liver and pancreas) and Sturge-Weber syndrome (facial and meningeal angiomas).

A special form of angioma of the skin, the so-called sclerosing angioma, is dealt with later (p. 1084).

Glomangioma (glomus tumour). This uncommon but interesting lesion apparently arises from the glomus bodies, small arteriovenous anastomoses with a coiled arteriole and abundant nerve supply

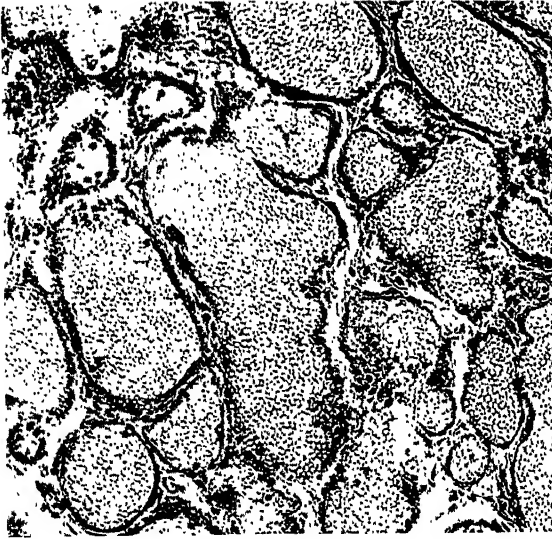


Fig. 13.17 Cavernous angioma of subcutaneous tissue, showing large intercommunicating spaces filled with blood. $\times 130$.

which control blood flow and temperature, particularly in the fingers and toes. In its most characteristic form, the glomangioma is a small bluish nodule, usually near the end of a finger, and extraordinarily tender to even light touch. On microscopic examination the tumour is found to consist of two kinds of tissue variously interblended (Fig. 13.18). The first is angiomatous, with spaces containing blood, lined by endothelium, and separated by connective tissue containing varying amounts of smooth muscle. The other is cellular, with rounded or cuboidal cells

called 'myoid', as transitions to smooth muscle fibres can be found. The growth contains numerous medullated and non-medullated nerve fibres and the pain is apparently due to distensile pressures in the blood-containing spaces, though the painfulness is not in proportion to the neural content. A small dermal leiomyoma may likewise be painful and the two may be related in origin. Glomangiomas have been described in deeper tissues, including the gut, but the characteristic pain occurs only with those in the limbs.

Chemodectoma. Because of their close anatomical relationship with blood vessels it is convenient to consider here the tumours arising from the chemoreceptor organs, viz. the carotid body, glomus jugulare, organ of Zuckerkandl and no doubt other less clearly defined structures such as the aortic bodies. These tumours have also been called **non-chromaffin paragangliomas**; they do not appear to produce any endocrine effects.

Chemodectomas are usually benign, but their anatomical sites may render complete surgical removal difficult. Thus carotid body tumours, which are the commonest variety, closely embrace the bifurcation of the common carotid artery, and the glomus jugulare tumours involve the middle ear and present as recurrent bleeding aural polyps; they may also present intracranially. Microscopically the architectural pattern is similar to the tissue of origin, consisting of many small masses of cells of variable size, sometimes enclosed in a boxlike framework of fine fibrous tissue (Fig. 13.19). The tumour cells are usually polygonal and may be spindle-shaped in places but aberrant types with hyperchromatic nuclei are not

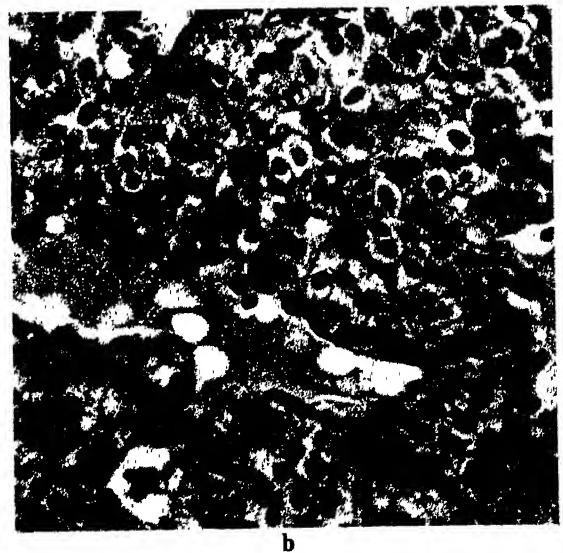
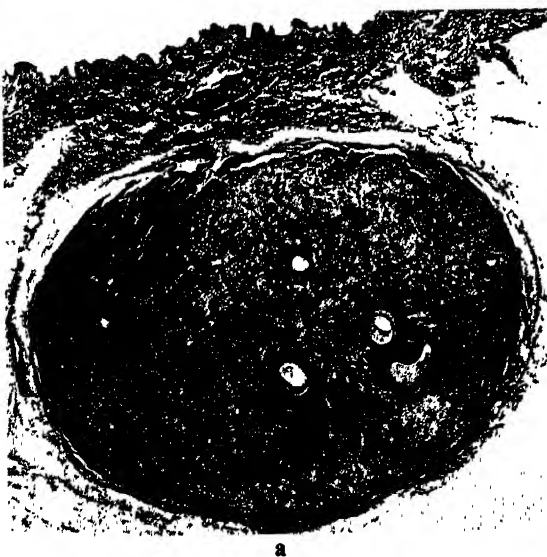


Fig. 13.18 Glomangioma.

- a** Small subcutaneous encapsulated growth showing the coiled arteriole. $\times 8$.
- b** The clear myoid cells surrounding a vascular space. $\times 350$.

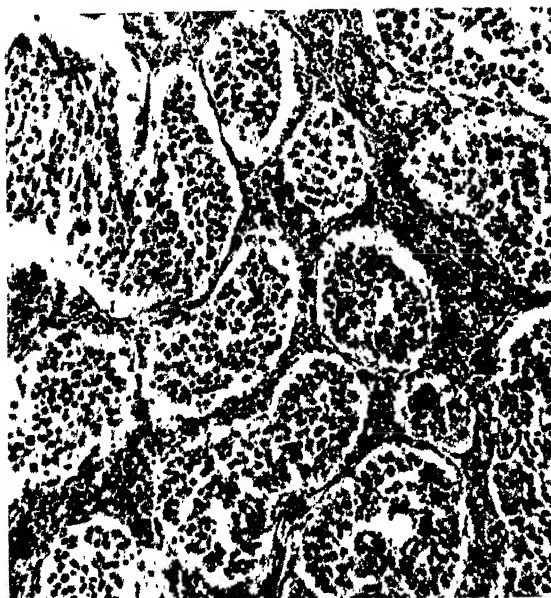


Fig. 13.19 Chemodectoma of the carotid body, showing the characteristic boxlike pattern and highly vascular stroma. $\times 160$.

uncommon and do not indicate malignancy. The blood supply is very rich and of sinusoidal pattern.

Malignant vascular tumours

With the possible exception of Kaposi's sarcoma (see below) which is relatively common in Central Africa but is of uncertain origin, true tumours of blood vessels are rare.

Occasionally a lesion with the histological features of haemangioma grows unusually rapidly and is classed by some as a *haemangio-endothelioma*. Very rarely an apparent capillary haemangioma gives rise to metastases and thus merits the term *haemangio-endotheliosarcoma*. This latter term is also applied to malignant tumours which show the cellular abnormalities characteristic of malignancy, but in which the neoplastic cells also show, in places, the structural arrangement of endothelial cells lining a lumen.

Rare tumours also arise from vascular pericytes (*haemangio-pericytoma*), the tumour cells being separated from vascular endothelium by basement-membrane material demonstrable by silver impregnation staining techniques. Some remain localised, but others metastasise.

Exposure to vinyl chloride monomer predisposes to the otherwise very rare *haemangiosarcoma of the liver* (p. 312), but it is not certain that this tumour arises from vascular endothelium.

Kaposi's sarcoma. This remarkable disorder is rare in Britain. It is more frequent in some other parts of Europe and Africa, and relatively common in some

well-defined areas of Central Africa, where it is much more frequent in males. The condition presents the syndrome of lymphoedema, multiple cutaneous tumours which later ulcerate, lymphadenopathy and ultimately visceral involvement. The skin tumours at first consist of lobulated masses of highly vascular and cellular tissue resembling granulation tissue deep within the corium and separated by fibrous trabecula (Fig. 13.20). Later there is much haemorrhage in and around the lesions, the spindle cells and vascular sprouts increase progressively, mitoses are abundant, ulceration of superficial lesions occurs and the regional lymph nodes may be replaced by similar highly vascular spindle-celled tissue. Despite their tumour-like appearance many lesions ultimately stop growing and heal with scarring, although the disease may progress elsewhere. However, in many cases (the proportion is uncertain), the lesions become progressive in internal organs, especially the intestines. This occurs more frequently in Central African cases than in those in South Africa or elsewhere. The superficial lesions respond well to radiotherapy and chemotherapy. The condition is sometimes associated with lymphomas.

It is uncertain whether this is a true neoplasm but in some cases it certainly behaves like one. Evidence from histochemistry and tissue culture indicates that the spindle cells are not fibroblasts: the lesion may



Fig. 13.20 Kaposi's sarcoma. An early lesion showing the zone of fibrous tissue between the vascular spindle-celled tumour and the epidermis. $\times 130$.

be an angiosarcoma derived from lymphatic endothelium as the cells lack the enzymes characteristic of blood capillary endothelium.

Tumours of lymphatics

Lymphangioma. This may be composed of numerous lymphatic vessels—the *plexiform* lymphangioma—but more frequently it has a *cavernous* structure. Dilatation and diffuse growth of vessels may give rise to enlargement of a part, e.g. the tongue (*macroglossia*). In such lesions there is even less evidence than in haemangiomas of neoplastic growth, and the more diffuse lesions may be hard to distinguish from the effects of lymphatic obstruction, though in most cases it is clear from the anatomy that no such obstruction can be present, and a congenital malformation (or hamartoma—p. 358) of the lymphatics is present. It is becoming increasingly common to describe such lesions as *lymphangiectasis* rather than lymphangioma. Lesions, whether diffuse or compact, are commonest in the skin and subcutaneous tissue. Each forms a somewhat ill-defined, doughy or semi-fluctuant swelling, containing large, intercommunicating lymphatic spaces. They contain clear lymph with occasional lymphocytes. Sometimes bleeding into the spaces renders the diagnosis between haemangioma and lymphangioma difficult. Lymphangiomas occur occasionally also in mucous membranes in the wall of the bowel (Fig. 13.21), in the tissues of the orbit and mesentery, and elsewhere.

In rare cases lymphangiomas of neck, retroperitoneum or mesentery undergo great dilata-



Fig. 13.21 Cavernous lymphangioma of small intestine. It is made up of large intercommunicating spaces filled with clear lymph (which has coagulated and then shrunk in processing the tissue). It occupies both mucosa and submucosa, the mucosa being much distorted but otherwise not much damaged.

tion, forming a multilocular ramifying cystic mass which may become very large. Occasionally a single cyst is formed which may be distinguished from other cysts by its endothelial lining.

Lymphangio-endothelioma (lymphangiosarcoma) is a very doubtful entity, but has been described as arising in the lymphatics of the arm, following their obstruction by mastectomy and irradiation for mammary cancer. Very similar appearances may result from a slow, diffuse permeation of the lymphatics by carcinoma and it remains undecided whether this is the real nature of the so-called lymphangiosarcoma.

Tumours of Neuro-Ectodermal Origin

From the ectodermal cells of the neural tube and crest are derived the tumours of (a) the neuroglia, (b) nerve cells and their precursors, (c) nerve sheaths, (d) peripheral neuro-receptor organs, (e) the melanocytes, and probably also of (f) the meninges. These tumours will be discussed in detail in the chapters on the nervous system and the skin. It will suffice at this stage to mention very briefly a few types and their features.

Tumours of neuroglia

These tumours, termed the **gliomas**, are the commonest primary tumours of the central nervous system. They occur mostly in the brain, but also rarely in the spinal cord. Because of their anatomical site and the pressure effects they may exercise on vital structures most gliomas eventually kill the patient. Even those of high cellularity and aberrant cellular structure,

which often spread widely within the cranial and spinal cavities, do not metastasise in extracranial tissues except rarely after surgical treatment involving craniotomy. (This rarity of distant metastases applies to all intracranial tumours, and appears to be a characteristic of the intracranial site, not of the type of tumour.)

Gliomas may take origin from all types of neuroglial cells and may be slow-growing, and rich in glial fibrils, but may nevertheless undergo central necrosis and become cystic. Even these well-differentiated tumours have ill-defined margins and differ from most benign tumours in having no capsule. At the other extreme are highly pleomorphic rapidly growing cellular tumours in which necrosis and haemorrhage are conspicuous.

Tumours of nerve cells

Perhaps because mature neurones are permanent cells which do not divide, nerve cell tumours are very rare in adults.

Melanocytic tumours

The cells which form melanin are of neural crest origin and are called melanocytes (p. 1078). The skin and the eye are their chief sites, and in both places they are important sources of tumours.

Skin tumours are dealt with fully in Chapter 27. The so-called **pigmented naevus** is so common that few people are free of them. It is a benign lesion, better regarded as a hamartoma than a true tumour, formed by a mass of melanocytes ('naevus cells') which accumulate in the dermis as a result of excessive proliferation of melanocytes in their normal site in the basal layer. **Malignant melanomas** are much less common, but common enough to be an important type of cancer. They arise from epidermal melanocytes, often at the site of a pigmented naevus. They are highly malignant and, unless excised at an early stage, are liable to metastasise extensively, both to lymph nodes

Medulloblastoma, a rapidly growing tumour of primitive nerve cell type, occurs in children, usually in the cerebellum (p. 787).

Ganglioneuromas are rare tumours composed of mature ganglion cells and nerve fibres. They arise chiefly in the sympathetic chain and adrenal medulla (p. 789). Clinically and histologically they behave as benign tumours except when they contain foci of neuroblasts.

Neuroblastoma or **sympathicoblastoma** is a highly cellular tumour which occurs chiefly in the adrenal medulla and in the sympathetic chain, usually in children. It consists of small round or oval cells with scanty cytoplasm, in places arranged in ball-like clusters which, on section, appear as rosettes. These tumours are highly malignant and metastasise widely, especially to the skull and other bones (p. 788).

Although strikingly different in appearance, it appears that ganglioneuroma and neuroblastoma are benign and malignant variants of the same tumour. Some differentiation commonly occurs in a neuroblastoma. Rarely, it matures completely into a ganglioneuroma.

and via the blood: since the tumours are often very dark due to melanin production, they can present a striking picture at necropsy.

Ocular melanomas. Malignant melanomas of the eye, though far from common, are the least rare of all intraocular malignant tumours (p. 802).

Melanomas at other sites. Melanocytes spill over all the mucocutaneous junctions into the adjoining mucous membranes to varying distances and in varying numbers, and melanomas occur at the corresponding sites. They are least rare in the nose, but arise also in the mouth, conjunctiva, vagina and anus and even in such deeper sites as oesophagus and rectum. The presence of melanocytes in the meninges is also reflected in the rare occurrence of meningeal melanomas, which, like other intracranial tumours, are remarkable for their inability to metastasise outside the cranial cavity.

Tumours of the Haemopoietic, Lymphoid and Mononuclear Phagocyte Systems

Classification within this group is difficult, and it is only in the last few decades that some of its most important members—the leukaemias and Hodgkin's disease for instance—have become fully accepted as neoplastic. Nomenclature still tends to be anomalous and classification disputed. The principal varieties are described in Chapters 17 and 18 and only a few general points will be made here.

The cells of these three systems are all derived from pluripotent haemopoietic stem cells which are present in the haemopoietic tissue and in small numbers in the blood (p. 116). Another common feature is the continuous production of the various cell types throughout life. Thirdly, the mature cells of these systems leave their site of origin in the haemopoietic marrow and circulate in the blood. Lymphocytes (which are produced also in the thymus) multiply in the secondary lymphoid tissues in response to antigenic stimuli and recirculate continuously between the blood, secondary lymphoid and other tissues. Monocytes settle as macrophages in nearly all tissues in the body, but especially in the spleen and lymphoid tissues, the liver, lung and haemopoietic marrow.

Immature precursors of cells of the blood do not normally escape into the blood, but may do so in various conditions of stress, and the spleen and liver can revert to their fetal role of haemopoiesis.

With all this movement and wide distribution of the normal cells of these systems, it is not surprising that neoplastic cells arising from them often do not form a single tumour mass, but spread widely from the start. The more slowly-growing, better differentiated neoplasms tend to spread particularly to the sites which normally contain the corresponding type of non-neoplastic cells. For instance, in chronic myeloid leukaemia, a well-differentiated neoplasm of the granulocyte series, very large numbers of tumour cells appear in the blood from the start, but the solid organs most involved are usually the marrow, liver and spleen, i.e.

the normal sites of haemopoiesis at various stages of life. By contrast, the poorly differentiated acute leukaemias produce far fewer mobile cells in the blood (and sometimes none at all) and tend to a more destructive infiltration of the marrow and other organs. Similarly the better differentiated lymph node tumours characteristically cause widespread moderate enlargement of multiple lymph nodes and other lymphoid structures but often do not spread very much to non-lymphoid tissues, while the less well differentiated may cause large local masses and are less discriminating in spread to other structures.

Any tissue which contains any of the cells belonging to the relevant categories (which means practically any tissue in the body) can give rise to tumours within this group, but the vast majority arise either in the bone marrow (leukaemias and myeloma*) or the lymph nodes (lymphomas, including Hodgkin's disease). As will be obvious from what has been said above, a benign tumour in the ordinary sense of the word must be exceptional, although non-progressive monoclonal proliferation of plasma cells is common in old age. Localised nodules of abnormal lymphoid tissue are sometimes found, for example in the rectum and skin, and may be called 'benign lymphoma' or 'benign lymphocytoma' but it is very doubtful whether these are tumours at all.

The immunological importance of these tissues is naturally reflected in their tumours. Immunological abnormalities are fairly commonly associated with some lymphoid tumours, the most striking example being production of immunoglobulin in large amounts by the neoplastic plasma cells of myeloma.

The viral aetiology of lymphomas and leukaemias in several animal species is now beyond doubt, and it is very likely that these human tumours are also virus-induced: so far, however, the evidence is largely indirect, except in the special case of Burkitt's lymphoma (p. 304).

*Myeloma is a neoplastic proliferation of plasma cells, and thus a B-cell lymphoma.

'Mixed' Tumours

A considerable number of tumours consist of two or more different kinds of tissue. The reasons for this are very diverse: the most important can be classified as follows:

(a) '*Collision*' tumours result when two different tumours arise close together and intermingle; the juxtaposition may be a chance event, or the result of a local carcinogenic stimulus affecting several tissues.

cinomas ('*adeno-acanthomas*'), with mixed squamous and glandular elements, which arise especially in the uterus and the bronchi.

(d) *Tumours with variable differentiation*. This occurs especially in carcinomas, parts of which are so poorly differentiated as to resemble sarcomas. It is almost certain that most so-called '*carcinosarcomas*', both human and experimental, are of this type (Fig. 13.22).



a



b

Fig. 13.22 Squamous carcinoma of tongue, showing so-called carcinosarcoma.

a On the left typical squamous carcinoma, showing continuity with spindle-shaped epithelial cells.

b On the right a purely spindle-cell area containing many mitoses. Transitions from (a) to (b) are readily found. $\times 200$.

(b) *Stromal changes in epithelial tumours*. The cartilaginous metaplasia of the stroma in mixed-salivary tumours, the bony metaplasia seen in a very few carcinomas of the colon, and the lymphoid stroma of some seminomas all produce the *appearance* of a mixed tumour, though only the epithelial element is truly neoplastic.

(c) '*Metaplastic*' tumours. Tumours arising from tissues which readily undergo metaplasia from one type to another often reflect this characteristic. Such are the adenosquamous car-

(e) With the exception of teratoma (see below) *true mixed tumours* involving the coordinated neoplastic growth of two independent tissues are hard to find. The fibroadenoma of the breast is the only completely acceptable example, though some connective tissue tumours, such as the angiolipoma, may qualify.

(f) '*Embryonal*' tumours. There is a group of tumours, mostly highly malignant, which arise usually in infancy and appear to be derived from immature tissue. Most of these (neuroblastoma, medulloblastoma, hepatoblastoma, for

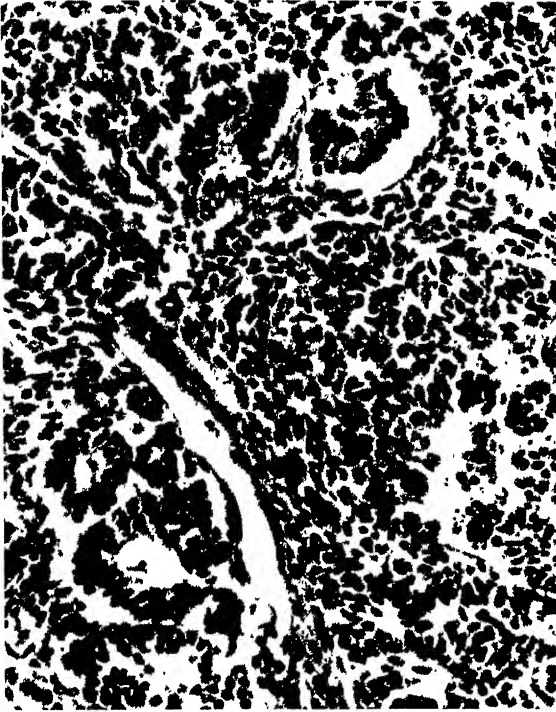


Fig. 13.23 Nephroblastoma (Wilms's tumour), showing cellular sarcoma-like tissue with indistinct differentiation of tubules in several places and a glomerulus-like structure (at top right). $\times 225$.

example) are not truly mixed, but the partial differentiation of some of the elements, while others continue to resemble primitive embryonal tissue, may give rise to appearances that simulate a mixed tumour.

One of the commonest of these tumours, the renal nephroblastoma (Wilms's tumour), consists of sarcoma-like masses of short spindle cells, among which some elements differentiate to tubules and occasionally glomerulus-like structures (Fig. 13.23). More discordant tissues such as cartilage and striped muscle are occasionally present, possibly representing a derivation of the tumour cells from the myotome at an earlier stage.

(g) *Teratomas*. These, the most extreme examples of mixed tumours, are dealt with in the following section.

It should be emphasised that the grouping together of these tumours is not intended to indicate any special relation between them: they have nothing in common except the presence of more than one kind of tissue.

Teratoma and Choriocarcinoma

As stated above, *teratoma* is the outstanding example of a true mixed tumour; it is of sufficient importance and interest to consider more fully. *Choriocarcinoma* is not a mixed tumour: it is described here because it sometimes develops within a teratoma, but it also occurs alone, originating usually from placental trophoblast.

Teratoma

A teratoma is a tumour composed of various tissues, chaotically arranged and usually of the most diverse types, with no relation to the site of origin. They are not rare and are of practical importance. They are most common in the ovaries and testes, though they occur in other parts, such as the mediastinum, retroperitoneal tissues and pineal. They are usually single but occasionally more than one is present. There is great variation in naked-eye appearances, and

cyst formation may be a notable feature, as in the common benign ovarian teratoma ('dermoid'). There is endless variety in the tissues and in their arrangement. Cartilage, bone, epidermis, glandular epithelium, hair, teeth, etc., are common components, especially in the benign forms, but other specialised tissues e.g. hepatic, renal, nervous, ocular and haemopoietic are also found (Figs. 13.24, 13.25). While a teratoma may be of such a complicated constitution, there is no proper formation of organs, limbs, etc., and a very important fact is that there is no trace of a vertebral column and no metameric segmentation. Germ cells and germinal epithelium are also always absent.

Because of their complicated structure, teratomas were formerly believed to arise from totipotent cells, i.e. from dislocated blastomeres. However, proliferation of a blastomere gives rise to an organised embryo, in



Fig. 13.24 Benign ovarian cystic teratoma or 'dermoid'. Above there is a cleft lined by squamous epithelium which is part of the main cyst. The cyst on the left is lined partly by columnar, partly by non-keratinising squamous epithelium. The lowest cyst is lined by folded columnar epithelium of alimentary type, and the resemblance to gut is heightened by an incomplete layer of smooth muscle related to it. $\times 72$.

contrast to the chaotic mixture of tissues in a teratoma. Another view is that teratomas are derived from the male or female germ cells. The frequency of teratomas in the gonads supports this possibility. Something of the nature of parthenogenetic development would have to be assumed, as is known to occur in amphibian ova under the influence of certain salt solutions. In cocks the intratesticular injection of solutions of zinc salts during the breeding season or after stimulation by pituitary gonadotrophin has led to the development of highly malignant complex teratomatous tumours closely resembling those in man (Bagg, 1936) but similar results do not appear to have been achieved in mammals.

Study of the sex chromatin (p. 1002) has provided some evidence in favour of the parthenogenetic origin of teratomas. It has been found that all teratomas in women have nuclei of the

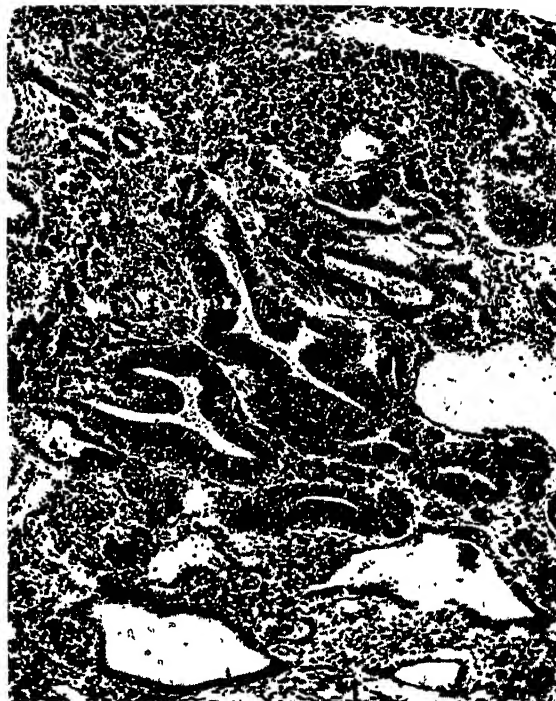


Fig. 13.25 Testicular teratoma (malignant). Because of loss of differentiation the diversity of tissue types is harder to demonstrate, but there is obviously a variety of epithelia present here, the large dark-staining tubules being of primitive neural type. $\times 62$.

right sex (i.e. are XX) but about half of those in men have female nuclei, as would be expected if they were produced by fusion of two haploid cells in the male gonad. Recent evidence from chromosome studies supports this view: the genes carried in the cells of a benign teratoma are identical with those of the host, but are redistributed between the members of each pair of chromosomes in a manner which normally occurs only during meiosis.

Teratomas are commonest in the ovary, in which site they are usually benign: elsewhere they are nearly always malignant. One type of tissue in a benign teratoma sometimes undergoes malignant change and when this happens the metastases consist of that one tissue only. In malignant teratomas of the type seen especially in the testis, all the tissues present are malignant from the start and metastases may contain any or all of them.

Benign teratomas never contain tissues of extra-fetal origin, such as trophoblast (see below) or yolk sac: their presence in a tumour, either alone or as part of a teratoma, always indicates malignancy.

Monstrosities. Any sufficiently striking congenital structural abnormality may be called a monstrosity. When identical twins develop from a single fertilised ovum a variable amount of fusion may take place. This may be of limited extent as in so-called Siamese twins, or it may affect a considerable part of the body. Partial fusion of two germinal areas is often invoked. Then there are cases where one fetus is imperfectly represented and fused with the other, growing on it in parasite-like fashion or included within its abdominal cavity—*fetus in fetu*. Many sacral teratomas and epignathi probably belong to this group. All such abnormalities, which are extremely varied, are spoken of as monstrosities. They show wide deviations from the normal, but the formation of parts and the relations of the tissues to one another are well maintained: thus, organs may be doubled, and a limb, though abnormal in size or form, is still a limb. Monstrosities are now generally thought to be a fundamentally different nature from teratomas and apparent transitions are almost certainly fallacious. A defect in the mechanism of the primary organisers at a very early stage of formation of the embryo may well be responsible for this type of abnormality.

Choriocarcinoma

This is uncommon except in the Far East. It is a highly malignant tumour which usually originates from, and retains recognisable features of, trophoblastic epithelium (Fig. 13.26). It also retains the trophoblastic property of invading blood vessels, and so metastases are usually early and widespread. Choriocarcinoma arises occasionally from trophoblastic differentiation in a teratoma of the testis, mediastinum, etc., but its usual site is in the uterus, where it origi-



Fig. 13.26 Choriocarcinoma, showing Langhans' cells (cytotrophoblast, *bottom right*) and giant cells (syncytiotrophoblast, *centre*) invading uterine wall (*left*). $\times 230$.

nates from the placental trophoblast of a pregnancy and is thus a fetal tumour growing in the mother. In spite of this, it grows and extends rapidly, but responds unusually well to chemotherapy, and in some instances removal of the primary tumour has been followed by apparently spontaneous disappearance of metastases. These unusual features, which apply only to the uterine (fetal) choriocarcinomas, are probably due to a homograft reaction.

Tumour-like Lesions and Cysts

There are a number of lesions which resemble tumours but have distinctive features which cast doubt on their neoplastic nature. Examples already described in this chapter include the fibromatoses (p. 342), haemangiomas (p. 348)

and monstrosities. Among such tumour-like lesions the groups termed *hamartomas* merit further description. *Cysts* are not tumours and are described here for want of a better place.

Hamartoma

This is a convenient term for an ill-defined group of lesions which have some resemblance to tumours but are not neoplastic. They usually appear before or soon after birth, grow with the individual and cease to grow when general body growth ceases. They may consist of a single type of cell, e.g. pigmented naevi, composed of a collection of melanocytes (p. 1080), a particular type of tissue, e.g. haemangioma, or a mixture of tissues, e.g. cartilage, epithelial-lined clefts, adipose and fibrous tissue in the so-called *adenochondroma* of the lung (p. 496). Those in the internal organs form lumps which can be mistaken grossly for tumours, while in the skin they may be nodular or may present as a patch of discolouration, e.g. the capillary haemangioma (p. 348) and some pigmented naevi. Hamartomas can best be understood as arising from a localised disorder of the relationships of normal tissues leading to overproduction of one or more elements but without the property of progressive growth characteristic of tumours. There are many varieties, and some have a tendency to progress to true neoplasia, for example pigmented naevi (although the risk is small) and exostoses (bony projections with a cartilaginous cap) arising from the axial skeleton or proximal limb bones (p. 904).

Cysts

The term 'cyst' properly means a space containing fluid and lined by epithelial cells. In most cysts the epithelial lining is not neoplastic and such cysts are neither tumours nor parts of tumours: they are included here only for convenience. Nearly all cysts arise by the abnormal dilatation of pre-existing tubules, ducts or cavities, though a cyst may lose its cell lining due to inflammatory or other change, and come to be lined by granulation or denser fibrous tissue. The term is, however, often applied in a somewhat loose way to other abnormal cavities containing fluid. For example, the term 'apoplectic cyst' is applied to a space in the brain containing brownish fluid, which has resulted from haemorrhage. Some tumours, e.g. gliomas, undergo softening in their interior, so that a collection of fluid is formed, and the term 'cystic change' is often used even when no true cyst is formed.

The cysts peculiar to each organ will be described in the later chapters: we shall give here only a classification of their causes. True cysts also occur in some tumours, the lining epithelium being neoplastic, e.g. in *cystic adenomas* (p. 326) and *teratomas* (p. 355). Apart from these, cysts fall naturally into two main groups: (1) those due to congenital abnormalities, and (2) acquired cysts, i.e. those produced by lesions in post-natal life.

(1) Congenital cysts

These may also be grouped into two types:

(a) They may arise *within otherwise normal organs or tissues*, as a result of the presence of epithelium of a type not usually present at that site after birth, either as a result of some minor displacement of an embryonal tissue or (more often) the failure to disappear of some embryonic duct or cleft. The commonest site of cysts derived from vestigial ducts is the genito-urinary tract, where the disappearance of the mesonephros and its duct in both sexes, and of the Wolffian ducts in females and Müllerian ducts in males, often leaves behind a variety of persistent epithelial remnants: small cysts are very common among these, and larger ones (*parovarian cysts*) are not uncommon in the broad ligaments (p. 970).

Other embryonic ducts which may persist and give rise to cysts include the thyroglossal duct (mid-line of neck, usually near the hyoid, p. 1030) and the urachus (usually at the umbilicus). A similar mechanism operates with the branchial clefts; **branchial cysts** are produced at the side of the neck and are lined by squamous epithelium with usually a rim of lymphoid tissue.

A different mechanism produces the **sequestration dermoids** which result from imperfect fusion of embryonal skin flaps. They are lined with squamous epithelium and filled with keratin, and are found mostly in the mid-line of the chest and neck or at the angles of the eye.

The 'pearly tumour' of the meninges, etc. (actually a squamous-epithelium-lined cyst, p. 792) is an example of a simple displacement of squamous epithelium into the meninges, at the time of neural tube closure.

(b) Cysts arising as *part of a major congenital abnormality of an organ*. Examples are (i) **polycystic disease of the kidneys** (p. 860), in

which a major maldevelopment of the renal tubules (of several possible types) results in the formation of cysts in great numbers; (ii) the **meningocele** and other types of cystic swelling that complicate some cases of spina bifida, failure of proper closure of the neural tube being the basic defect (p. 772).

(2) Acquired cysts

These are of several varieties, the three following being the most important:

(a) **Retention cysts.** These are formed by retention of secretion produced by obstruction to the outflow of secretion. A single cyst, sometimes large, may be produced by the obstruction of the main duct, e.g. of a salivary gland or of a part of the pancreas. Obstruction of the orifice of a hair follicle gives rise to a cyst-like swelling filled chiefly with breaking-down keratin—the so-called **sebaceous cyst**, seen especially in the scalp. On the other hand, numerous small cysts may result from obstruction of small ducts, an occurrence which is not uncommon in fibrosing lesions of the kidney.

(b) **Distension cysts** are formed from natural enclosed spaces. They occur in the thyroid from dilatation of the acini, and occasionally also in the pituitary: cystic dilatation of Graafian follicles in the ovaries is also common. Distension of spaces lined by mesothelium is also seen; for example, a bursa may enlarge to form a cystic swelling, and there is the common condition of hydrocele due to an accumulation of fluid in the tunica vaginalis.

Occasionally in the adult an **implantation cyst** occurs by the dislocation inwards of a portion of epidermis by injury. The epithelium grows and comes to line a space filled with degenerate epithelial squames (Fig. 13.27); rarely hair follicles are present in the wall. Implantation cysts may result also from wounds of the cornea.

(c) **Parasitic cysts.** These are cystic stages in the life cycle of cestode *parasites*. The most striking examples are the 'hydatid' cysts produced in man, usually in the liver, by the dog tapeworm *Taenia echinococcus*, though small cysts may be produced in the brain and other parts by the cysticerci of *Taenia solium*.

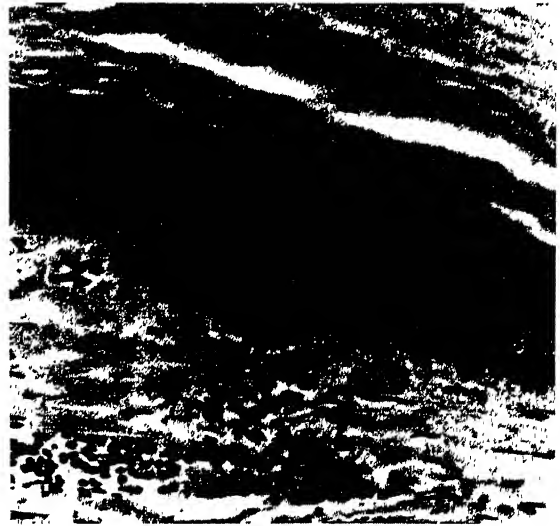


Fig. 13.27 Implantation cyst in subcutaneous tissue, showing lining of stratified squamous epithelium and keratin in the lumen. $\times 200$.

Further Reading

- Ashley, D. B. (1978). *Evans' Histological Appearances of Tumours*, 3rd edn., pp. 857. Churchill-Livingstone, Edinburgh, London and New York. (An account of the behaviour and appearances of human tumours based on a considerable experience.)
- Sobin, L. H., Thomas, L. B., Percy, Constance and Henson, D. E. (Eds.) (1978) *A Coded Compendium of the International Histological Classification of Tumours*, pp. 116. World Health Organisation, Geneva. (A widely accepted system of classification, coding and nomenclature of human tumours.)
- Willis, R. A. (1973). *The Spread of Tumours in the Human Body*, 3rd edn., pp. 417. Butterworths, London.
- Atlas of Tumour Pathology*. US Armed Forces Institute of Pathology, Washington, DC. (Numerous 'Fascicles' on tumours of particular organs, tissues and regions. A valuable source of detailed information on the histology and behaviour of individual tumours.)
- Cancer*. A journal of the American Cancer Society. Lipincott, Philadelphia and Toronto. (A monthly publication of well-illustrated articles on human neoplasms.)

Blood Vessels and Lymphatics

Arteries

Introduction

Lesions of the arteries are very important because of their frequency and serious consequences. The commonest important disease in developed countries is **atheroma (atherosclerosis)**, which consists of prolonged, slow, patchy accumulation of lipids and fibrosis, in the intima of arteries of various sizes. The patchy thickening results in narrowing of the lumen with consequent chronic ischaemia of the various organs and tissues. Acute ischaemia can result from the occlusion of an artery by local **thrombosis** or **embolism**; these processes have been dealt with in Chapter 9, but it is important to emphasise here that *arterial thrombosis is usually the result of disease of the artery wall, and atheroma is the commonest predisposing cause.*

Another very common arterial change is termed **arteriosclerosis**. This is a diffuse change in the walls of arteries, in which the muscle and elastic tissue slowly diminish and are replaced by fibrous tissue, with the result that the arteries become more firm and rigid. Like atheroma, these changes occur in various degrees with increasing age, but they are aggravated and accelerated by systemic hypertension in which the fibrous replacement is preceded, at least in some instances, by hypertrophy of muscle and elastic tissue. Arteriosclerosis alone is usually without serious consequences, but the accompanying changes in the arterioles—**arteriolosclerosis**—result in ischaemia, particularly of the kidneys.

A very common and important condition which affects the arteries and arterioles is **systemic hypertension**, a state in which the arterial blood pressure is raised. In most cases, the cause of the rise in blood pressure is not

known, but it is likely that, in all instances, the rise is mediated by increased muscle tone in the arterioles. Prolonged hypertension, as mentioned above, leads to severe arteriosclerosis and arteriolosclerosis; it is by far the commonest cause of rupture of the cerebral arteries to produce haemorrhage into the brain, and is an important cause of heart failure.

Diseases of the arteries which bring about severe destruction of the muscle and elastic tissue, particularly of the media, may weaken the wall to such an extent that dilatation results, and if localised this is termed an **aneurysm**; rupture of the vessel wall may occur, with or without preceding dilatation. Severe weakening can result from various forms of **arteritis**, including that due to syphilis, but also from **atheroma** and from *degenerative changes of unknown nature in the media*. All three of these conditions can give rise to aortic aneurysm. Aneurysm formation and rupture may result also from *developmental defects*, as seen in the arteries at the base of the brain. Another effect of the various types of arteritis is to promote thrombosis, although atheroma is a much more important cause of this in certain arteries.

Effects of ageing. Throughout adult life, the walls of arteries of all sizes become gradually less resilient and more rigid; they tend to enlarge both in diameter and in length. These changes constitute **senile arteriosclerosis**: they are well illustrated by the prominence and tortuosity of the temporal arteries in older people, and are due to gradual increase in collagen and ground substance at the expense of smooth muscle and fine elastic fibres. The media is mainly affected, but also the intima which becomes appreciably thickened. Chemical analysis has shown that there is also a gradual in-

crease in calcium salts in the artery walls. Senile arteriosclerosis has little effect on function. Similar changes, but with some thickening of the artery walls, are a feature of chronic hypertension.

Patchy thickening and hyalinisation of the walls of arterioles (**hyaline arteriosclerosis**) also occurs with increasing age, particularly in the spleen and kidneys, but also in other viscera. When severe, it results in luminal narrowing, and may thus cause ischaemia, e.g. of the renal glomeruli. Hyaline arteriosclerosis is exaggerated in hypertension and diabetes: it is described more fully on p. 373.

Adaptation and hypertrophy. The muscular and elastic nature of the arterial wall readily permits dilatation to provide an increased blood flow to the part supplied. If the requirement is more than transient, the lumen becomes persistently dilated and there is hypertrophy of the muscular and elastic tissue, e.g. the physiological hypertrophy of the uterine arteries during pregnancy. Similar compensatory dilatation and hypertrophy occurs in collateral vessels when a main artery is obstructed.

In persistent hypertension, the tendency for the increased pressure to dilate and lengthen the arteries is partly prevented by compensatory hypertrophy of the circular muscle of the media and the intimal longitudinal muscle fibres which lie next to the internal elastic lamina (Fig. 14.1). The vascular changes in hypertension are described in more detail on pp. 371–4.

Endarteritis obliterans. Intimal thickening of arteries occurs when concentric laminae of cellular connective tissue form in the intima and obliterate or narrow the arterial lumen. New elastic tissue is laid down independently of the internal elastic lamina and may appear as a layer under the endothelium or as a number of small new laminae. This lesion is called **endarteritis obliterans** and is found in chronic inflammatory lesions. It is well seen in the base of chronic peptic ulcers (Fig. 14.2) and in the walls of tuberculous cavities, where it is beneficial in tending to prevent haemorrhage. It occurs also in the lesions of syphilis, in tuberculous meningitis (Fig. 21.33, p. 751) and silicosis, and in small arteries exposed to radiotherapy. Similar obliterative changes occur in the smaller arteries in severe hypertension and in progressive systemic sclerosis.



Fig. 14.1 Longitudinal section of the wall of an artery in essential hypertension, showing hypertrophic thickening of the longitudinal muscle in the intima *a* and of the circular muscle in the media *b*. (Myocytes stained black.) $\times 110$.

When the functional requirements for blood flow through an artery are greatly reduced, the lumen is narrowed by obliterative endarteritis, which is the physiological mechanism of arterial involution. It occurs, for example, in the umbilical arteries and ductus arteriosus after birth, in the uterine and ovarian arteries after the menopause, and in the arteries supplying an area which is excised (e.g. a limb) or destroyed by disease.

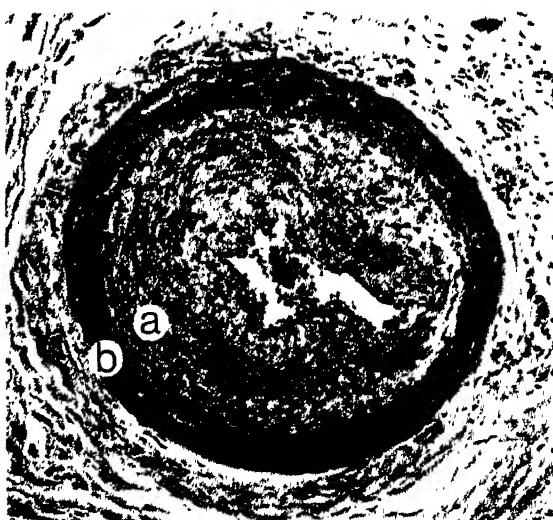


Fig. 14.2 Endarteritis obliterans in the base of a chronic peptic ulcer, *a* = intima, *b* = media.

Atheroma (Atherosclerosis)

This is one of the most important diseases of developed communities. It causes narrowing of the lumen of arteries, is often complicated by occlusive thrombosis, and is the major cause of disability and death from heart disease, cerebral infarction and ischaemia of the lower limbs. It is virtually always present in some degree in middle-aged and old people.

Definition. The lesions of atheroma consist of patches ('plaques') of intimal thickening of the walls of arteries, due mainly to deposition of lipids and formation of fibrous tissue. The alternative term **atherosclerosis** is used because the lesion has a soft, lipid-rich part (athere = porridge) and a hard (sclerotic) fibrous component.

Naked-eye appearances. The earliest deposits of lipid in the intima of the aorta and large arteries are seen predominantly in childhood and adolescence and are known as *fatty streaks*. They appear as yellow non-raised spots in the luminal surface, which enlarge and coalesce to form irregular yellow streaks. Microscopy shows them to consist of accumulations of lipid droplets in intimal cells (now known to be smooth muscle cells—see below) and in aggregates of macrophages lying beneath the endothelium (Fig. 14.3). Fatty streaks are seen in children dying from various causes, including trauma, and are apparently equally common in all communities, regardless of whether or not there is a high incidence of atheroma later in life. Because they are found mainly in children, many of the fatty streaks

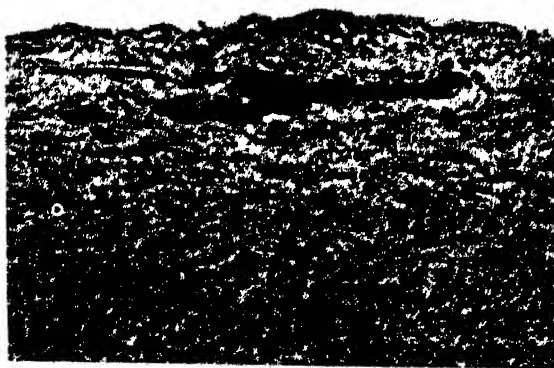


Fig. 14.3 Intimal lipid deposit in the aorta of a child aged 11 years: accidental death. (Frozen section; lipid stained black.) $\times 200$. (Dr. Morag McCallum.)

must disappear, but it is not known whether some of them persist and progress to atheroma in communities with a high incidence of this disease.

The earliest recognisable atheromatous lesions are seen commonly from young adulthood onwards in atheroma-prone communities. They consist of small disc-like yellowish, slightly raised patches of intimal thickening with a smooth glistening surface. As the condition progresses, the patches enlarge and thicken by further deposition of lipid deep in the intima and by fibrosis more superficially (i.e. adjacent to the lumen): the patches become distinctly raised and when viewed from the intimal surface they may appear yellow or white depending on the amount of white fibrous tissue overlying the yellow lipid deposits. In any one individual, the patches are at various stages of development, indicating progressive formation.

Aorta. Atheroma occurs throughout the length of the aorta, but the abdominal aorta is usually more severely affected and patches often develop first around the origins of the intercostal and lumbar branches (Fig. 14.4). The



Fig. 14.4 Mild atheroma of the abdominal aorta. The lesions are seen as raised patches and are located mainly around the origins of the arterial branches. $\times 0.8$.



Fig. 14.5 Lengths of the abdominal aorta: *left*, minimal atheroma; *middle*, severe atheroma with cracking and early ulceration of patches; *right*, very severe atheroma with ulceration and mural thrombosis. Note also that the two atheromatous aortas have lost their elasticity and stretched: this may be due to atrophy of the media beneath the extensive atheroma, but could also be the result of arteriosclerosis.

patches vary in size up to several centimetres diameter and may in places become confluent. If a sizeable patch is cut across, lipid-rich paste-like material can be expressed from its deeper part, and fibrous thickening is seen as a white layer overlying this. The fibrous layer may break down, resulting in *ulceration* of the plaque, and *mural thrombus* is then likely to be deposited on the ulcerated surface; another common change is *deposition of calcium salts* which may convert the plaque to a hard brittle plate. Plaques showing ulceration, calcification or thrombus deposition are commonly referred to as **complicated atheroma**, and may produce great irregularity of the luminal surface of the aorta (Fig. 14.5). Other important features of aortic plaques are *thinning of the overlying media*, and in some instances *extension of the plaque into the adjacent media*: the wall is thus weakened and an *aneurysm* may develop, with the danger of rupture (p. 385).

Other arteries. Atheroma occurs in arteries of all sizes down to approximately 2 mm diameter and is seen occasionally in mild form in even smaller vessels. The general features are similar to those seen in the aorta except that the plaques are necessarily smaller, often involve the whole circumference of the intima, and can cause all degrees of *luminal narrowing* down to virtual occlusion (Figs. 14.6, 14.7,



Fig. 14.6 Atheroma of a coronary artery causing moderate narrowing of the lumen. The spaces in the deep part of the plaque represent lipid accumulation. $\times 10$.

14.8). Atheroma tends to affect especially arteries supplying the heart, brain and abdominal viscera, and also the arteries of the lower limbs. There is considerable individual variation in its distribution, in some instances the aorta being mainly affected, in others the arteries at the base of the brain and/or the coronary arteries. The coronary arteries are often severely affected and are more often involved at a relatively early age than any other arteries.

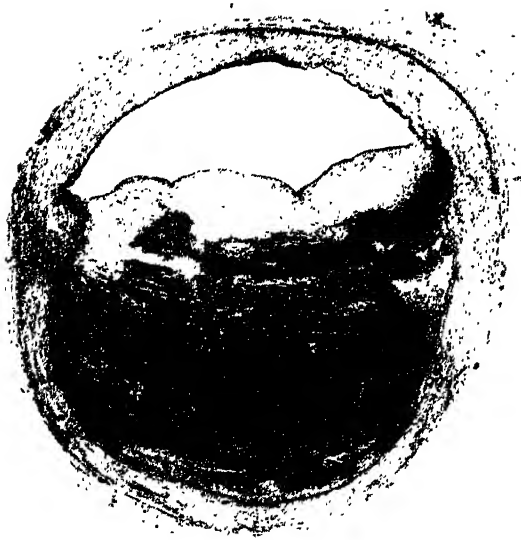


Fig. 14.7 Severe atheroma of the superior mesenteric artery, causing marked reduction of the lumen. Frozen section, stained with Scharlach R, showing the large amount of fatty material in the patch. $\times 15$.



Fig. 14.8 Severe atheroma of the left coronary artery in a patient with myxoedema. The lumen has been greatly reduced by atheroma and occluded by recent superadded thrombosis. $\times 10$.

The cerebral arteries also are subject to severe atheroma, but this is found chiefly in elderly persons. For unknown reasons, the renal arteries are seldom severely affected except in diabetics, in whom atheroma is often widespread and very severe.

In the thin-walled arteries at the base of the brain the patches are visible from both inner and outer aspects of the vessels, and their yellow opaque appearance contrasts with the reddish translucency of the normal parts of the vessel wall (Fig. 14.9).

Complications include: (a) **haemorrhage into a plaque**, which increases the degree of luminal narrowing (Fig. 15.2, p. 401); (b) **rupture or ulceration of a plaque** (Fig. 15.12, p. 406); and (c) **occlusive thrombosis** (Fig. 15.6, p. 403) which is a major cause of infarction in the heart, brain and intestine, and of ischaemia of the legs.

Microscopic appearances. The early changes are due to accumulation of lipids in proliferated spindle cells, shown by electron microscopy to be smooth muscle cells, lying in the intima (Fig. 14.10). Lipids also accumulate between cells deep in the intima (i.e. close to the media), particularly in relation to elastic fibres



Fig. 14.9 Circle of Willis and branches, showing marked patchy atheroma.

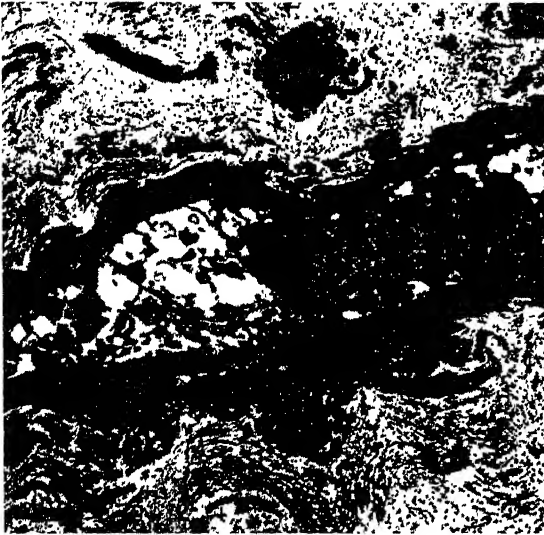


Fig. 14.10 Electron micrograph of part of a smooth muscle cell in an atheromatous plaque. The cytoplasm is made up largely of myofibrils, and contains globules of fat, shown as light spaces. The fine fibres on either side of the cell are collagen. $\times 7500$.

and the internal elastic lamina. As the patch develops, thin laminae of connective tissue appear in the more superficial part of the intima and form the fibrous part of the lesion. Lipid-containing cells lie among the collagen fibres in this region. Areas of necrosis then develop, converting the deep part of the patch into a structureless accumulation of lipids, tissue debris (Fig. 14.11) and sometimes altered blood, and the necrosis gradually extends into the overlying fibrous tissue. Calcium deposition may be visible microscopically. Infiltration of neutrophil leukocytes and other inflammatory cells is common, and lipid-laden macrophages—'foamy cells'—may appear around the lipid deposits, which usually contain crystals of cholesterol, represented in paraffin section by the typical elongated clefts (Fig. 14.11). The internal elastic lamina deep to the plaque is usually disrupted and lipid deposition, necrosis and fibrosis may then extend into the adjacent media. Quite apart from this, the media deep to the plaque becomes thinned and atrophic (Fig. 14.12).

Small blood vessels grow into the atheromatous patch from the media of the affected vessels and sometimes also from the intimal surface. These may be the source of the haemorrhage which commonly occurs in the patch, although, as already stated, the overlying fib-

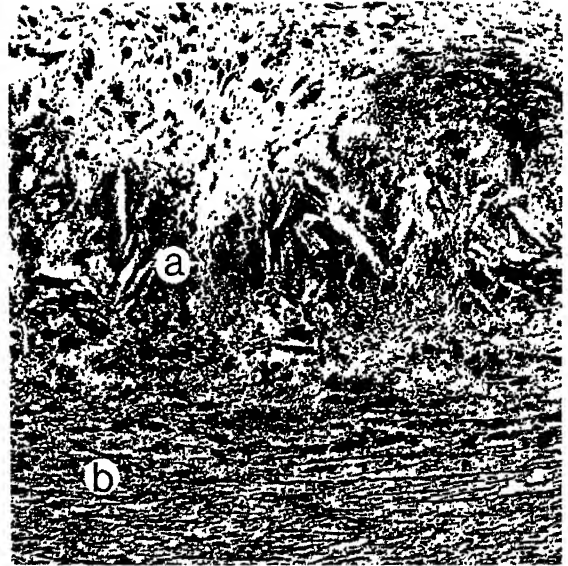


Fig. 14.11 Section showing part of an atheromatous patch of the aorta. In the deep part of the intima there is degenerate lipid-rich material **a**, the spindle-shaped spaces being due to cholesterol crystals. Lipid accumulation stops abruptly at the junction with the media **b**. $\times 110$.

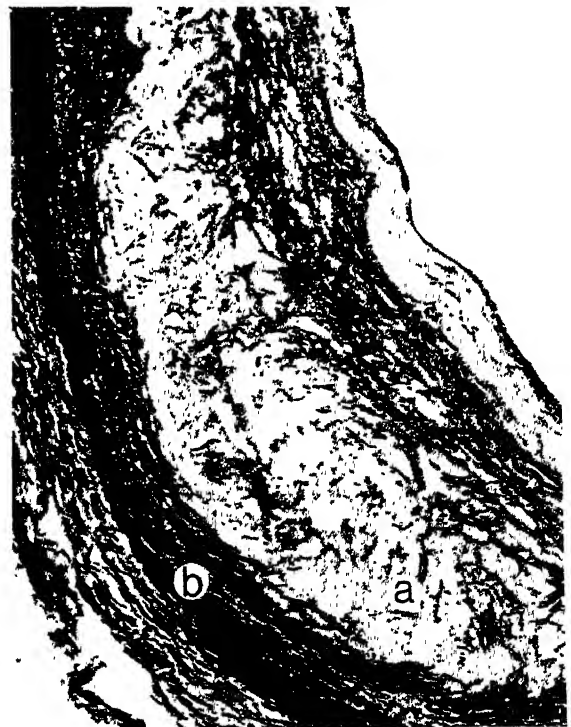


Fig. 14.12 Atheroma of a cerebral artery. The intima is greatly thickened with accumulation of lipid in its deeper part (**a**) and dense overlying fibrosis. The media (**b**) shows local atrophy over the patch. $\times 110$.

rous patch may be very thin and may rupture, allowing blood to track in from the lumen.

Effects

Although the changes of atheroma are essentially the same in all arteries, their effects vary in arteries of different sizes.

Large arteries. Uncomplicated atheroma of large arteries, such as the aorta, very often has no clinical effect because usually it does not substantially reduce the lumen or seriously weaken the wall. In advanced cases, however, an *aneurysm* may form in the abdominal aorta (p. 385) or occasionally in a common iliac artery. *Thrombi* which form on ulcerated plaques in the aorta seldom cause complete occlusion, probably because the rapid flow limits platelet adhesion. Occasionally, however, thrombus may extend to occlude the whole aortic lumen and when this occurs at the bifurcation it can result in gangrene of the legs unless adequate collateral circulation has developed, in which case there may be merely coldness and weakness of the legs with muscle wasting and sexual impotence but without ischaemic pain or gangrene (*Leriche syndrome*). Apart from occluding the aorta, thrombi and atheromatous debris from ulcerated plaques may break away and form *emboli* in the arteries of the lower limbs and abdominal organs such as the kidneys.

Smaller arteries. By far the commonest important effects of atheroma are due to involvement of smaller arteries, the lumen of which may be progressively narrowed by an atheromatous patch or suddenly occluded by super-added thrombosis (Fig. 14.8). These effects are well seen in the coronary arteries. Atheroma is the chief cause of *ischaemic heart disease*, the largest single cause of death in Europe and North America today (p. 400). *Ischaemic brain damage* is also very common and is usually the result of atheroma of the carotid, vertebral and basilar arteries, vessels of the Circle of Willis and cerebral arteries (p. 742). Atheroma does not cause aneurysms of smaller arteries.

Arteries supplying the legs are often severely atheromatous, with consequent progressive diminution in blood supply. Eventually the collateral circulation becomes inadequate: relative **muscle ischaemia** can then be induced by the **increased metabolic demands** of exercise, which

produces severe pain in the leg, relieved by rest. This is the clinical syndrome of *intermittent claudication*. In time ischaemia may be so severe as to cause *gangrene*, which usually starts in the toes (Fig. 2.6, p. 12) and spreads proximally. Examination of legs amputated for gangrene usually shows narrowing or obliteration and calcification of the main arteries of the leg. Because atheroma is often widespread, patients with severe involvement of the arteries of the lower limbs frequently suffer also from ischaemic heart disease.

The arteries of the arms are rarely severely affected by atheroma.

Aetiology

Epidemiological surveys have revealed a number of predisposing factors in atheroma, but the mechanisms involved in the development and growth of the plaque are not yet understood. Most of the available information has been provided by morphological, histochemical and biochemical studies on the atheromatous plaque, and by experimental animal studies. It seems appropriate to discuss these biomedical investigations, and then to consider how the various known risk factors may contribute.

The main features of the early lesion are accumulation of lipid, cellular proliferation and formation of fibrous tissue.

Accumulation of lipid. Studies in man with labelled cholesterol suggest that most of the cholesterol in the human aortic intima is derived from the plasma. Since the arterial vasa vasorum supply only the outer part of the wall, plasma lipid must enter the intima from the lumen via the endothelial lining, and animal studies have shown that plasma proteins and certain lipids normally pass into the intima via this route, particularly around the mouths of arterial branches. This localisation suggests increased endothelial permeability at sites where atheroma tends to occur in man, and where lesions resembling human atheroma can be induced experimentally by feeding animals on a cholesterol-rich diet. Such focal increase in permeability is associated with an increased rate of endothelial cell-turnover and it has been suggested that the endothelium at certain sites, including the vicinity of arterial branching, is subject to increased shearing stress which may

account for its greater permeability and turnover rate. There is no good evidence on the route of transport of lipid across the endothelial barrier: micropinocytic vesicles or the development of gaps between cells (p. 49) are obvious possibilities. Other agents which may play a causal role in atheroma have been shown to cause endothelial-cell injury, for example inhalation of cigarette smoke by rabbits results in focal endothelial-cell loss, while the formation of circulating antigen-antibody complexes, which increase endothelial permeability by inducing gaps between endothelial cells (p. 154), has been shown to enhance the dietary-induced atheroma-like lesions of rabbits. In man, cardiac allotransplants tend to develop extensive and heavy deposition of lipids in the intima of the coronary arteries, possibly as a result of endothelial injury resulting from the reaction of host antibody with donor vascular endothelium.

Cholesterol is transported in the plasma as a component of lipoproteins (p. 24), which are now usually classified into *chylomicrons* and three other groups which differ in their composition, electrophoretic mobility and specific gravity—the α or *high density lipoproteins* (HDL), the β or *low density lipoproteins* (LDL) and the *pre- β or very low density lipoproteins* (VLDL). The LDL are particularly rich in cholesterol, much of which is in the form of cholesterol esters rich in linoleic acid, and a high proportion of the cholesterol in early atheromatous plaques is also in this form. When pieces of aorta are incubated in culture medium containing lipoproteins, both HDL and LDL (but not VLDL) pass through the endothelium, but LDL accumulate in the intima in much greater amounts than HDL: there appears to be some mechanism of clearing HDL from the intima which is not effective for LDL. Immunological assay of lipoproteins in the human aortic intima, and particularly in early atheromatous patches, has demonstrated disproportionately large amounts of LDL.

Findings such as those outlined above have led to the widespread belief that the lipid of atheromatous plaques is derived from the plasma, that LDL make a major contribution to it, and that injury to the endothelium probably plays an important role in allowing lipid to enter the intima.

Cellular proliferation. The intimal cells which proliferate in early atheroma, and which accu-

mulate droplets of lipid, have been shown by electron microscopy to be modified smooth muscle cells (Fig. 14.10) which migrate from the media, through the internal elastic lamina, into the intima. Two factors have been suggested to account for their proliferation. Firstly, it has been shown that LDL promote proliferation of smooth muscle cells in aortic explants in culture. (Curiously, only LDL prepared from plasma in which the LDL level is high appear to have this effect.) Secondly, when platelets adhere to an injured vessel wall or form aggregates, they discharge their storage granules (the platelet release reaction—p. 232), and one of the stored products released is a basic protein which stimulates proliferation of smooth muscle cells in culture.

So far, we have considered accumulation of lipid in the intima by the process of insudation from the plasma (**filtration theory**), but with mention of platelets the time is opportune to introduce the **thrombogenic theory** of atheroma, which was suggested by Rokitsansky in 1852, only to be forgotten and re-introduced independently by Duguid in 1946. This theory proposes that injury to arterial endothelium results in recurrent deposition of a fine layer of mural thrombus consisting of platelets and fibrin, and that the thrombus is rapidly covered by endothelium and thus incorporated into the superficial part of the intima. It is further suggested that such thrombus is the origin of the lipid of atheroma and that its recurrent deposition accounts for the gradual development and growth of the atheromatous plaque.

By use of labelled antibodies, both fibrin and platelets have been detected in approximately 40 per cent of atheromatous patches. Experimentally, injury to aortic endothelium, e.g. by abrasion, has been shown to result in adherence of platelets which soon become covered by endothelium: platelet deposition is followed by active proliferation of smooth muscle cells in the intima, with formation of a fibromuscular plaque resembling early atheroma. Depression of platelet adherence by dipyridamole or anti-platelet antibody inhibits the proliferation of smooth muscle cells in such experiments.

It may be concluded that deposition of a thin layer of platelets and fibrin on the surface of atheromatous patches is a common occurrence, and that such thrombus is rapidly incorporated

into the plaque. The role of thrombus in the development and growth of the plaque is uncertain, and many workers consider that it is unlikely to be the major source of atheromatous lipid.

The fibrous tissue which forms the superficial (sub-endothelial) part of the atheromatous plaque is provided, at least in part, by the proliferated smooth muscle cells, which have been shown in tissue culture to be capable of producing collagen, elastin and the proteoglycans of ground-substance. These are curious cells, reminiscent in their properties of the myofibroblasts of healing wounds (p. 84). Another surprising observation has been provided by the study of iso-enzymic forms of glucose-6-phosphate dehydrogenase. These enzymes are encoded by X chromosomes, and many American negroes are heterozygous, so that, as a result of Lyonisation (p. 513), half of their cells produce one type (A) of iso-enzyme and the remainder produce another (B). Analysis of early cellular atheromatous plaques from such heterozygous women has shown that the smooth muscle cells of some plaques produce only iso-enzyme A, while those of other plaques (from the same individual) produce only B. This suggests that the cellular proliferation in each plaque is *monoclonal*, although there are other possible explanations. This remarkable finding may be of basic importance in atherogenesis (see Benditt, 1977).

The necrosis which occurs in the lipid-rich depth of the atheromatous plaque may be due to ischaemia, although the deeper part of the plaque is supplied by the vasa vasorum which extend into the (normally avascular) intima of affected arteries. Another possible factor is the presence in the plaque of small amounts of unusual lipids shown to be capable of inducing tissue injury.

The story of focal endothelial injury resulting in ingress of lipids by insudation and in deposition of platelets seems a reasonable explanation of the development of early atheroma. It does not, however, account for the growth of the plaque once a well-defined superficial layer of fibrous tissue has formed: neither insudation from the lumen nor mural thrombosis would account for further increase in lipid deep in the intimal plaque, the most likely source of which seems to be the vasa vasorum.

Predisposing or 'risk' factors

While the pathogenesis of atheroma is not fully understood, there are a number of factors which are known to predispose to its development. These are as follows.

Blood lipids. An important relationship between blood lipids and atheroma is indicated by the following observations.

- (1) There is a positive correlation between the average level of the serum or plasma cholesterol (which depends mainly on diet) in a community and the incidence and severity of atheroma.
- (2) Individuals with a high plasma total cholesterol level have an increased risk of atheroma, and the incidence of severe atheroma is very high in subjects with diseases accompanied by hypercholesterolaemia, for example diabetes mellitus, myxoedema, familial hyperbetalipoproteinaemia (p. 30) and the nephrotic syndrome.
- (3) Lesions resembling atheroma can be produced in various animal species by a cholesterol-rich diet.

Hyperlipidaemia is now classified into the various types of hyperlipoproteinaemia, which differ in the type of lipoprotein which is increased and in their dependence on genetic and dietary factors. It is obviously important to determine whether atheroma is related to rise of a particular lipoprotein fraction and, if so, whether the hyperlipoproteinaemia can be corrected by diet. Unfortunately there is no simple method of assessing the presence, distribution and severity of atheroma during life unless it gives rise to symptoms. Accordingly, it is usual to regard ischaemic heart disease (IHD) as an indication of severe atheroma, and to seek associations between possible risk factors, e.g. hyperlipidaemia and IHD. It must be appreciated that IHD is a crude indication of atheroma, for in some patients with IHD, atheroma is limited mainly to the coronary arteries, and many people without IHD have severe atheroma. Also, IHD is commonly due to occlusive coronary artery thrombosis superimposed on atheroma, and accordingly it is not known whether risk factors implicated by such studies predispose to atheroma, to superadded thrombosis, or to both.

Nevertheless, useful information has been provided by population studies. In the Framingham investigation in Massachusetts, re-

ported by Kannel (1971), the levels of serum cholesterol, phospholipid, LDL and VLDL were determined in over 5000 individuals, who were then followed up for 16 years. Ischaemic heart disease, indicated by myocardial infarction, sudden death or angina pectoris, was shown to correlate strongly with each of these lipid fractions. Since this study was undertaken a classification or hyperlipoproteinaemias has been published by the World Health Organisation (1970). Extensive prospective studies on the relationship between the types of hyperlipoproteinaemia and atheroma have not yet been reported, but it is already apparent that there is a greatly increased risk of atheroma in individuals with high levels of LDL and VLDL. The two commonest types of hyperlipoproteinaemia (IIb and IV), both of which carry a high risk of IHD, are attributable to dietary excess, either of animal fats or of total caloric intake, and *it seems very likely that the incidence and severity of atheroma, or at least of IHD, could be greatly diminished by modification of the diet.*

The association of LDL with a high risk of IHD fits well with the biomedical studies described earlier. Many cells, including smooth muscle cells, ingest LDL by micropinocytosis and make use of its cholesterol in the production of cell membranes. As mentioned on p. 30, the cell uptake of LDL is controlled by cell receptors, etc.: the subject is more fully explained by Goldstein and Brown (1977). Defects of the control mechanisms result in the very severe atheroma of the familial hyperlipidaemias. No such defects have been demonstrated in the great majority of individuals with severe atheroma, but high levels of LDL passing from the plasma into the intima of arteries might overwhelm the normal control mechanisms, resulting in accumulation of LDL. Unsaturated vegetable oils do not raise the LDL and are not atherogenic: they may even have a protective effect.

There also appears to be an *inverse* relationship between serum HDL levels and ischaemic heart disease. A possible explanation is provided by experiments mentioned earlier in which pieces of aorta were incubated in culture medium containing lipoproteins: the presence of HDL was reported to result in transfer of cholesterol from the intima to the culture medium.

Age and sex. It is not surprising that

atheroma and its complications increase with advancing age, for the plaques grow slowly. Men are more severely affected than women at all ages, and indeed severe extensive atheroma is much less common in women until after the menopause, when it progresses as in younger men. The sex difference may be due to oestrogens, which are known to influence lipid metabolism and reduce the total plasma cholesterol.

Hypertension. There is no doubt that hypertension is associated with an increased incidence and severity of atheroma. This has been established by necropsy studies. Moreover, in the Framingham and other prospective studies, a clear correlation has been demonstrated between the height of the blood pressure and the risk of IHD, particularly in older men. It is not known how hypertension contributes to atheroma; one possibility is that it might promote endothelial injury by increasing the shearing stress on the endothelium.

It is noteworthy that the pulmonary arteries, in which the pressure is low, are usually not affected by atheroma except in patients with pulmonary hypertension, e.g. in mitral stenosis.

Cigarette smoking. The incidence of IHD in cigarette smokers is at least double that in non-smokers, and the risk increases with the number of cigarettes smoked daily. The duration of the habit does not appear to be an important factor and this, together with the fact that sudden deaths from IHD are especially increased in smokers, suggests that smoking may promote coronary artery thrombosis rather than atheroma itself. As mentioned earlier, inhalation of cigarette smoke has been shown to cause vascular endothelial injury in rabbits. It has also been shown that plasma fibrinogen, clotting factor VIII and the haematocrit levels are all raised in chronic smokers and that nicotine stimulates release of nor-adrenaline and thus promotes vasoconstriction. The role of smoking in atheroma and IHD thus appears complex.

Physical activity. The Framingham and other studies have demonstrated beyond doubt that the incidence of IHD is lower in people who are physically active than in more sedentary individuals. Exercise may be protective by using up lipids and carbohydrates for energy production, and study of the blood lipids has demonstrated relatively high levels of HDL in

active subjects: as stated above, HDL appear to protect against atheroma.

Psychological factors. Emotional stress appears to predispose to IHD, perhaps by increasing the output of catecholamines. Psychological factors may also play an indirect role by influencing choice of occupation, smoking or eating habits.

Prospective studies have shown that the various risk factors are synergistic: for example, an individual who smokes and has a high blood cholesterol level, hypertension and obesity, is *especially* likely to develop IHD. The factors predisposing to IHD have been shown (e.g. by Kannel *et al.*, 1976) to contribute also, in various degrees, to the risk of developing cerebral infarction and ischaemia of the lower limbs, both of which are nearly always complications of atheroma.

Systemic hypertension

Definition

Blood pressure usually increases with age, but there is considerable individual variation in the increase, and recordings of the blood pressures in a general adult population show a wide range. Any definition of hypertension must therefore be arbitrary, and there is no general agreement on the level of blood pressure to be regarded as pathological. Indeed, there is good evidence, for example from insurance companies' statistics, of a general inverse relationship between the height of the blood pressure, including variations within the 'normal range', and the expectation of life.

Classification

In about 85 per cent of cases of hypertension the cause is not apparent and these patients are said to have **primary**, **essential** or **idiopathic** hypertension. In the remaining 15 per cent hypertension is **secondary** to other disease processes: nearly always diseases of the kidneys are responsible ('renal hypertension') but occasional cases result from certain functioning adrenal tumours or as a feature of Cushing's syndrome (see Table 14.1). Coarctation of the aorta (p. 428) is accompanied by hypertension in the arteries arising proximal to the constriction. It is likely that, as diagnostic techniques improve,

further causes of hypertension will be identified, and the proportion of patients with so called essential hypertension will thus become smaller. Conn's syndrome (primary hyperaldosteronism, p. 1041) is an example of a condition which has been distinguished relatively recently from essential hypertension.

Table 14.1 Classification of systemic hypertension

I. Essential	{ benign malignant
II. Secondary	{ benign malignant
(a) of renal origin ('renal hypertension'), due to:	
	chronic pyelonephritis
	glomerulonephritis
	diabetes
	polycystic disease of the kidneys
	renal amyloidosis
	connective tissue diseases, particularly poly- arteritis
	urinary tract obstruction (occasional cases)
	renal artery disease
	radiation nephritis
	some renal tumours
	some congenital diseases of kidney, possibly by predisposing to pyelonephritis
(b) adrenal-mediated hypertension.	
	Conn's syndrome (primary hyperaldosteronism)
	Cushing's syndrome
	phaeochromocytoma
(c) coarctation of the aorta.	

Regardless of the aetiology, hypertension may be divided into **chronic** or so-called '**benign**', and **malignant** (sometimes called **accelerated**) types. In benign hypertension the rise of blood pressure is usually moderate, although sometimes marked. Many patients with benign hypertension lead active lives for many years with few or no symptoms, and die of some independent disease. Unless the blood pressure is controlled by antihypertensive drugs, however, it frequently causes disability and death from heart failure, and also increases the risk of myocardial infarction and cerebral vascular accidents.

Malignant hypertension is characterised by a very high blood pressure, by eye changes which include retinal haemorrhages and exudates and sometimes papilloedema, by rapidly progressive renal injury terminating in uraemia,

and by hypertensive encephalopathy. The pathological hallmark of this state is fibrinoid necrosis of arterioles (see later). These special features appear to depend on the rapid development of a very high blood pressure. Unless treated, patients with malignant hypertension usually die within six months or so, but frequently the blood pressure can be reduced by anti-hypertensive drugs, and the outlook is then greatly improved.

Benign and malignant hypertension should not be regarded as unrelated conditions. Malignant hypertension supervenes in a small proportion of cases of benign essential hypertension although more often it arises apparently *de novo*, i.e. without evidence of preceding benign hypertension.

Changes in the blood vessels

Changes develop in arterial vessels of all sizes as a result of hypertension. In the larger arteries, from the aorta down to vessels of about 1 mm diameter, the changes are widespread, and are termed **hypertensive arteriosclerosis**. Changes in the vessels below this size, i.e. in the smallest arteries and arterioles, tend to affect especially the small vessels of the viscera, and in particular those of the kidneys. The changes occurring in the larger arteries are of the same nature in all types of hypertension, but those in the smaller vessels, particularly the arterioles, are different in benign and malignant types of hypertension, and require separate descriptions.

Large and middle-sized arteries. The vascular changes in hypertension, uncomplicated by the arterial lesions common in the aged, are most readily studied in young patients with high blood pressure secondary to renal disease. In the **early stages** they consist mainly of hypertrophy of smooth muscle and elastic fibres. In the aorta, there is increase in both of these elements in the media. In muscular arteries the increase is mainly in the circular muscle of the media (Fig. 14.13) but also in the longitudinal muscle fibres of the intima (Fig. 14.1, p. 361); the internal elastic lamina becomes thickened, and very often new laminae are formed towards the intima (Fig. 14.14). In **longstanding hypertension**, which is usually of benign essential type, these hypertrophic changes give way to fibrous replacement of muscle and the

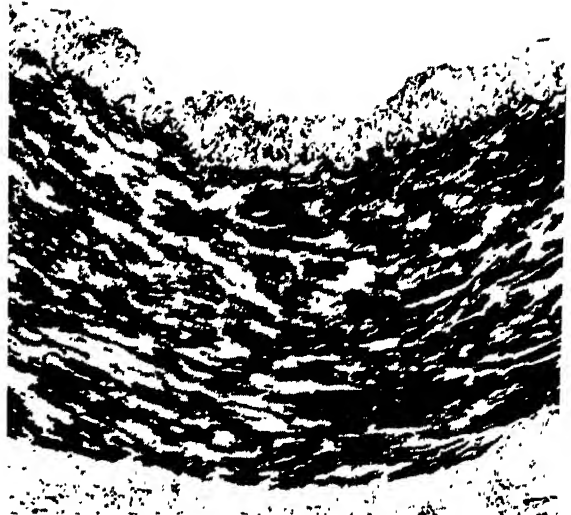


Fig. 14.13 Section of hypertrophied radial artery, from a case of chronic glomerulonephritis in a young subject, showing hypertrophy of the media. (Myocytes appear black.) $\times 140$.

elastic tissue may break up and undergo partial absorption. The arterial walls are thickened and of increased rigidity, the lumen is dilated (Fig. 14.15) and the vessels are often elongated and tortuous. In the aorta, there is increase in the elastic and fibrous tissue of the media. In the muscular arteries, the media is thickened and fibrosed with patchy loss of smooth muscle,



Fig. 14.14 Another section of the same artery as in Fig. 14.13, showing increase of elastic tissue formed by replication of the internal elastic lamina. (Elastic tissue appears black.) $\times 120$.



Fig. 14.15 Arteriosclerosis of the aorta and its branches in a patient with hypertension who died aged 36 from chronic renal failure. The walls of the vessels are thickened and rigid: they are also dilated, although this is not readily apparent. $\times 0.35$.



Fig. 14.16 Part of a transverse section of an arteriosclerotic artery. The intima (a) is thickened, while the muscle (shown as black) of the media (b) is partly replaced by fibrous tissue. (Compare with Fig. 14.13.) $\times 200$.

and there may be fibrous thickening of the intima (Fig. 14.16). These changes are widespread, and vary in degree. They are mostly without important effects. Hypertension increases the risk of rupture of the 'berry' aneurysms which develop in the arteries at the base of the brain in some individuals (p. 388), resulting in subarachnoid haemorrhage.

The arteriosclerotic changes described above are similar to those observed in normotensive elderly subjects (*senile arteriosclerosis*) but in the absence of hypertension they are usually less pronounced, and the media, although fibrosed, is often not thickened.

Atheroma tends to be particularly severe in individuals with chronic hypertension, and there is no doubt that prolonged elevation of the blood pressure aggravates this condition.

Hypertension thus results at first in hypertrophy of the arterial walls, with increase in muscle and elastic fibres, followed by arteriosclerosis and a tendency to severe atheroma. The early hypertrophic changes are usually observed only in young hypertensive subjects:

in older patients with chronic hypertension, arteriosclerosis and atheroma predominate.

Small arteries and arterioles. In arteries of 1 mm diameter or less, and in the arterioles, the changes differ from those in the larger vessels, and they differ also in benign and malignant hypertension.

(a) *Benign hypertension.* The small arteries show the medial thickening seen in the larger vessels, but a more pronounced degree of intimal thickening, due to concentric increase in connective tissue; in the smallest arteries the intimal change predominates, and may result in narrowing of the lumen in contrast to the dilatation seen in the larger arteries.

The arterioles undergo *hyaline thickening* of their walls (*hyaline arteriolosclerosis*), which consists at first of patchy deposition of hyaline material, often beneath the endothelium, but sometimes more peripherally: the hyaline change gradually extends to involve the whole circumference, and when severe it replaces the normal structures of the wall apart from the endothelium. This change occurs also apart

from hypertension, and is seen especially in old age. In both normotensive and hypertensive subjects it is observed most commonly in arterioles in the spleen, then in the afferent glomerular arterioles of the kidneys (Fig. 14.17), and in



Fig. 14.17 Hyaline arteriolosclerosis of afferent glomerular arteriole in chronic systemic hypertension. The arteriole is not only thickened, but also tortuous, and so has been cut twice in cross-section in the same plane. $\times 1500$.

the pancreas, liver and adrenal capsules. In all these sites, the change is appreciably commoner and usually more severe in hypertensives than in normotensive subjects of corresponding ages. Hyaline arteriolosclerosis is uncommon in the arterioles of the brain, gastro-intestinal tract, pituitary, thyroid, heart, skin and skeletal muscles (Smith, 1956). When severe, hyaline arteriolosclerosis results in considerable narrowing of the arteriolar lumen. This has important effects upon individual glomeruli, but does not usually cause renal failure. The nature of the change is not fully understood: initially, the hyaline material resembles fibrin in its staining properties, but later it stains like collagen. It also contains lipid material, and there is evidence that it may result from an exudative process in which plasma seeps into the arteriolar wall (p. 809). Apart from its occurrence as an ageing process and in hypertensive subjects, hyaline arteriolosclerosis is often severe and extensive in diabetes mellitus (pp. 840, 1033).

(b) *Malignant hypertension.* In this condition, the concentric fibrous thickening of the intima of the small arteries is often of extreme degree, particularly in the interlobular arteries of the kidneys (Fig. 22.10, p. 812). The arterioles are thickened and of hyaline appearance, as in benign hypertension, but the change is relatively acute, and consists of necrosis of the arteriolar wall, accompanied by permeation with plasma and deposition of fibrin (so-called **fibrinoid necrosis**): pyknotic nuclei, neutrophil polymorphs and red cells can often be found in the necrotic wall. The lumen is considerably narrowed, and superadded thrombosis may complete its occlusion. These changes affect especially the viscera, and the arteriolar lesions may result in haemorrhages and in ischaemic necrosis (Fig. 14.18). Focal fibrinoid necrosis may develop also in the small arteries.

In both benign and malignant hypertension, the changes in the small vessels in the kidneys cause renal damage, as described on pp. 809–12.

Cerebral haemorrhage in hypertensives is probably due to rupture of micro-aneurysms of the small arteries within the brain (p. 388).

Course of chronic and of accelerated hypertension

Chronic ('benign') essential hypertension. As already stated, this is much the commonest type of hypertension. The blood pressure rises very gradually over a period of years, in most cases to moderately high levels, e.g. 180/110 mm Hg, but occasionally much higher. The increase nearly always starts before the age of 45 years, and individuals with a resting blood pressure consistently below 140/85 at this age are very unlikely to develop essential hypertension. The diastolic pressure is less subject to physiological variations than the systolic pressure, and a diastolic pressure persistently exceeding 90 mm Hg is generally regarded, on an arbitrary basis, as abnormal. However, the disease develops very slowly, and it may be years before the rise in pressure clearly exceeds that which occurs normally with age.

The condition may be symptomless, and many cases come to light during routine medical examination for insurance or other purposes. *Common symptoms* include palpitations, audible pulsation in the head, headaches,

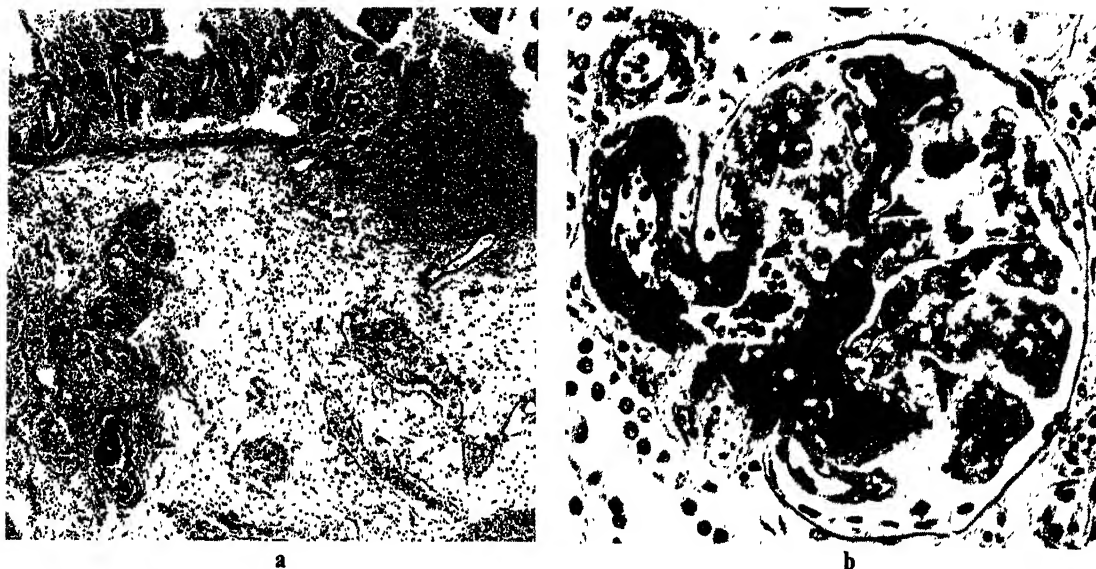


Fig. 14.18 Arteriolar lesions in malignant hypertension. **a** Ulceration of the colonic mucosa due to fibrinoid necrosis and thrombosis of arterioles: one such vessel is seen (*lower left*) in the submucosa. **b** Fibrinoid necrosis of a glomerular afferent arteriole and part of the tuft.

attacks of dizziness particularly on stooping, and reduced exercise tolerance.

About 60 per cent of deaths in those known to have 'benign' essential hypertension are from left ventricular or total heart failure; this is due to the increased work load thrown on the left ventricle and to the commonly associated severe coronary artery atheroma. About 30 per cent of patients die from cerebral haemorrhage, and the remaining 10 per cent from various causes unrelated to the hypertension. Although changes occur in the kidneys as a result of the vascular lesions, renal failure is uncommon in benign essential hypertension. When heart failure develops, however, there is usually a moderate rise in the blood urea level. In those patients who progress from chronic to malignant hypertension renal failure commonly supervenes.

Malignant ('accelerated') essential hypertension. This develops in approximately 10 per cent of cases of chronic essential hypertension. In those cases not preceded by chronic hypertension, the onset is usually between 30 and 45 years. It can result in heart failure or cerebral haemorrhage, but *without effective treatment renal injury is severe and usually causes death within a few months* (p. 811).

Eye changes are another important feature in malignant hypertension: lesions in the small arteries in the retina result in oedema, haem-

orrhages, infarcts and exudates, and blindness may ensue. Papilloedema, associated with cerebral oedema, is often present. *Hypertensive encephalopathy*, characterised by epileptiform fits and transient paralysis, is not uncommon, and is attributable to cerebral oedema, resulting from arterial spasm and focal cerebral ischaemia. This has been observed directly in rats with experimental hypertension and the fits have been shown to cease when the blood pressure is lowered and cerebral vasoconstriction ceases.

Secondary hypertension. Hypertension is a feature of **chronic renal failure**, and is more often of malignant type than is the case in essential hypertension. The superadded imposition of further renal injury from hypertensive vascular lesions, whether benign or malignant, aggravates and accelerates renal failure.

Aetiology of hypertension

In all types of hypertension, the raised blood pressure is a result of increased peripheral vascular resistance. In normal circumstances the peripheral resistance is controlled by the muscular tone in the arterioles throughout the body, and the major aetiological problem is the elucidation of the factors which, by increasing arteriolar tone, bring about the various types of hypertension. The possibility that the structural

changes of arteriolosclerosis initiate the hypertensive state is unlikely, for such changes are sometimes absent, particularly in early cases, and moreover structural changes do not develop in arterioles which are protected from hypertension by occlusive changes in the larger arteries supplying them. For these reasons, it is widely believed that the observed structural changes in the arteries and arterioles are the result of hypertension, and not the cause. It is likely, however, that structural changes in the arterioles and small arteries of the kidneys impair renal blood flow, and this may play a part in *maintaining* hypertension once the vascular changes have become pronounced.

Humoral vasoconstrictive factors. In patients with hypertension due to a pheochromocytoma, the large amounts of catecholamines released by the tumour into the blood (see p. 1047) are very likely to be the cause of the hypertension.

The vasoconstrictor substances renin and angiotensin (p. 258) are believed to be of importance in the pathogenesis of malignant hypertension secondary to renal disease: plasma concentrations of both renin and angiotensin are increased in this disorder. By contrast, their levels in most patients with benign essential hypertension are quite normal, and in patients with Conn's syndrome (primary aldosteronism, p. 1041) plasma renin concentration is actually reduced. Nevertheless, it cannot be concluded from this that renin and angiotensin do not contribute to the increased blood pressure in these conditions, as sensitivity to the pressor effects of injected angiotensin is known to be increased in hypertensive patients, and thus the normal or subnormal amounts of angiotensin in the blood might conceivably raise the blood pressure to abnormal levels.

Search for other vasoactive substances in the blood of patients with hypertension has failed.

Renal hypertension. A firm experimental basis for renal hypertension was provided in 1934 by Goldblatt and his colleagues, who showed that partial clamping of the renal arteries produced hypertension in dogs. This has been confirmed repeatedly in several species and it has been shown that hypertension can be produced in the rat by partial clamping of one renal artery. Vascular hypertensive changes have been produced by this method, and it is of interest that they do not

affect the kidney which is protected by the clamp from hypertension. Experimental hypertension of short duration produced in this manner may be abolished by removing the clamp or excising the clamped kidney, but if the clamp has been left in place for some months, the hypertension persists in spite of these manipulations, because of arteriolar changes produced in the unclamped kidney.

Renal hypertension in man is similar in many ways to the experimental condition. The diseases which cause it are listed in Table 14.1 on p. 370 and are described in Chapter 22: because of their relatively high incidence, *chronic glomerulonephritis* and *chronic pyelonephritis* are the most important ones. Release of excess renin from the abnormal kidney or kidneys may be an important factor in producing hypertension in such diseases, but the original view that the hypertension is due simply to the pressor effect of excessive renin production by the ischaemic kidney is now known to be an oversimplification. High levels of renin have usually been found only in malignant hypertension with underlying renal disease, and even in these the importance of renin is not fully established.

Secondary hyperaldosteronism in hypertension. In some cases of severe hypertension, particularly those with malignant hypertension and/or renal disease, secondary hyperaldosteronism develops. Plasma renin concentration is invariably increased and this leads to stimulation of aldosterone secretion which in turn produces potassium depletion (see p. 1038). The condition is recognised usually by a decrease in the concentration of potassium, and often of sodium, in the plasma. It must be distinguished from primary hyperaldosteronism (Conn's syndrome) in which the hypertension and hypokalaemia are associated with *increased* sodium and *decreased* plasma renin concentration (see p. 1041).

Neural factors in the pathogenesis of hypertension. The possibility that neural factors may play a role in primary hypertension deserves consideration. There is evidence that in both human and experimental hypertension the threshold of the vascular receptors is elevated, so that abnormally high pressures are necessary to initiate neurogenic anti-pressor reflexes. It may also be that variations in sensitivity to pressor agents, possibly genetically determined, are involved.

Pulmonary hypertension

In contrast to systemic hypertension, a rise in blood pressure in the pulmonary arterial system is usually explicable on the basis of disease of the lungs, heart or major vessels. These causes, and the effects of pulmonary hypertension, are described on pp. 454–7.

Calcification of the media (Mönckeberg's sclerosis)

Definition. This is a degenerative disease of unknown cause characterised by dystrophic calcification (p. 287) in the media, especially common in the major arteries of the lower limbs in elderly people. It may also affect the arteries of the upper limbs, and less commonly visceral arteries.



Fig. 14.19 Calcification of media of iliac artery, showing transverse markings caused by confluent calcification. $\times 0.8$.

Naked-eye appearances. The affected vessels are generally dilated and show transverse bars of medial calcification due to deposition of calcium in the circular medial muscle layer (Fig. 14.19). At a later stage, lengths of the arteries may be converted into rigid tubes. There may be no noteworthy alteration of the intima though atheroma is sometimes present in addition.

Microscopy shows that the earliest change is hyaline degeneration of the muscle fibres and connective tissue, usually starting about the middle of the media (Fig. 14.20). Calcium salts are deposited first as fine granules, and confluent calcification follows. There may be little or no cellular reaction. Occasionally true bone may be formed in an area of calcification, and may even contain red marrow.

Aetiology. This is generally regarded as an exaggeration of the natural increase of calcium salts in the arteries with age. It sometimes



Fig. 14.20 Calcification of media. The calcified tissue is darkly stained. $\times 16$.

occurs earlier in arteriosclerotic vessels but in man is not intimately related to high blood pressure. A similar lesion has been produced in the aorta of rabbits by injections of adrenaline.

Effects. The radiological appearance is striking but the lumens of the arteries are seldom narrowed. Ischaemic effects, if present, are usually due to co-existing atheroma and its complications.

Inflammatory lesions (arteritis)

Syphilitic arteritis

Small arteries. Wherever syphilitic lesions occur there is intimal and adventitial fibrosis of small arteries associated with infiltration of lymphocytes and plasma cells (endarteritis and periarteritis). In some cases the change is diffuse and a number of the vessels show general thickening, while in others it is of a patchy or nodular type.

Effects. Reduction of the lumens of small arteries and arterioles due to endarteritis may contribute to the necrosis in gummas. The essential change in syphilitic mesaortitis is probably involvement of the small nutrient vasa vasorum of the aortic media. Effects purely attributable to ischaemia are seen in the brain in tertiary syphilis (meningovascular syphilis) due to endarteritis obliterans of cortical vessels (Fig. 21.36, p. 753) and this may be complicated by thrombosis leading to cerebral infarction at an early age.

Syphilitic mesaortitis. This is a common manifestation of tertiary syphilis and an important cause of death in this disease. It occurs in acquired and congenital syphilis but has become increasingly rare in most developed countries.

Naked-eye appearances were formerly readily studied in untreated young subjects before the onset of arteriosclerosis and atheroma. The first visible lesions are greyish-white translucent areas of thickening in the intima, with little tendency to degenerate. Later they extend and fuse, forming areas with wrinkled 'tree bark' appearance: the intima between appears healthy (Fig. 14.21). In places, contraction of the tissue may occur with formation of stellate scars. Localised depressions which are potential aneurysms may be seen. In older subjects, yellow patches of atheroma, which is often



Fig. 14.21 The thoracic aorta in syphilitic aortitis. The arch of the aorta is stretched, with localised bulgings, thickened intimal patches and irregular wrinkling and scarring. In this example the changes stop abruptly below the arch. $\times 0.5$.

severe, may be associated with the syphilitic lesions. Occasionally, on cutting through the wall of the aorta, gummatous necrosis may be seen extending inwards from the adventitia.

The part of the aorta immediately above the aortic valve is usually involved first and the aortic arch is by far the commonest and at times the only part of the aorta with syphilitic lesions. Syphilitic changes are usually limited to the thoracic aorta.

Microscopic appearances. The earliest change is periarteritis and endarteritis of the vasa vasorum in the adventitia (Fig. 14.22). These changes then extend into the aortic media, in which foci of cellular infiltration appear (Fig. 14.23), with new formation of thin-walled vessels. This leads to breaks or windows in the elastic tissue and muscle of the media, best seen in a section stained to show the elastic fibres (Fig. 14.24). The elastic tissue and muscle are replaced by fibrous tissue, but gummatous necrosis may also occur.



Fig. 14.22 Syphilitic aortitis, showing severe endarteritis of an arteriole in the adventitia of the aorta, and infiltration by plasma cells and lymphocytes around two small vessels (*right*). $\times 330$.

Dense fibrous thickening of the intima occurs over the lesions of the media (Fig. 14.24), and vasa vasorum may extend into these intimal patches which account for the pearly-white raised areas seen by naked eye.



Fig. 14.23 Section of syphilitic aorta, showing cellular accumulations around the small vessels in the media, with destruction of the laminae. $\times 160$.



Fig. 14.24 Syphilitic aortitis; elastic tissue appears black. The section shows part of a thickened intimal plaque (*upper right*) and irregularity, thinning and interruptions in the elastic tissue of the media, which has also lost most of its muscle and is grossly thinned; the paler tissue below is the adventitia and adjacent fatty tissue. $\times 10$.

Effects. *Aneurysm formation* is an important complication of syphilitic mesaortitis and is due to weakening of the vessel wall from loss of medial elastic and muscle tissue: the effects of aneurysm are described on p. 385.

Aortic incompetence. The dilatation of syphilitic aortitis may involve the root of the aorta, with consequent incompetence of the aortic valve. The cusps become stretched, thickened and distorted (p. 419).

Coronary artery narrowing due to involvement of their orifices by mesaortitis is now a rare cause of myocardial ischaemia.

Thromboangiitis obliterans (Buerger's disease)

Definition. Buerger's disease is an inflammatory condition of arteries and veins, with thrombosis, organisation and recanalisation of the affected vessels. It occurs almost exclusively in men, and affects mainly the vessels of the lower limbs, but sometimes also those of the upper limbs, giving rise to severe pain and progressive ischaemic changes.

Pathological changes. The early changes are not often available for histological examination. They consist of occlusion of the affected vessel by thrombus which contains foci of intense polymorph infiltration. The whole

thickness of the vessel wall is also infiltrated with polymorphs. These acute changes give way to chronic inflammation, and the thrombus is replaced by granulation tissue containing lymphocytes, macrophages and multinucleated giant cells (Figs. 14.25, 14.26). The inflammatory changes eventually subside, and although the original vascular lumen has been obliterated there is often a surprising degree of recanalisation. The inflammation of the vessel wall also progresses to a chronic stage but without the degree of disruption which occurs in polyarteritis nodosa (see below). Fibrosis extends into the surrounding connective tissue, and the burned-out lesions thus consist of recanalised vessels with thickened fibrosed walls, enclosed in fibrous tissue which may envelop and compress adjacent nerves and vessels (Fig. 14.25).

The lesions affect short lengths of the small and medium sized arteries and veins of the legs and feet, but seldom the larger vessels. Similarly, when the upper limbs are affected, the lesions are mainly in the vessels of the forearms and hands. The disease is chronic, acute lesions



Fig. 14.25 Thromboangiitis obliterans. Occlusion of the posterior tibial artery (*upper left*) and surrounding fibrosis extending around the adjacent veins and nerves. $\times 32$.

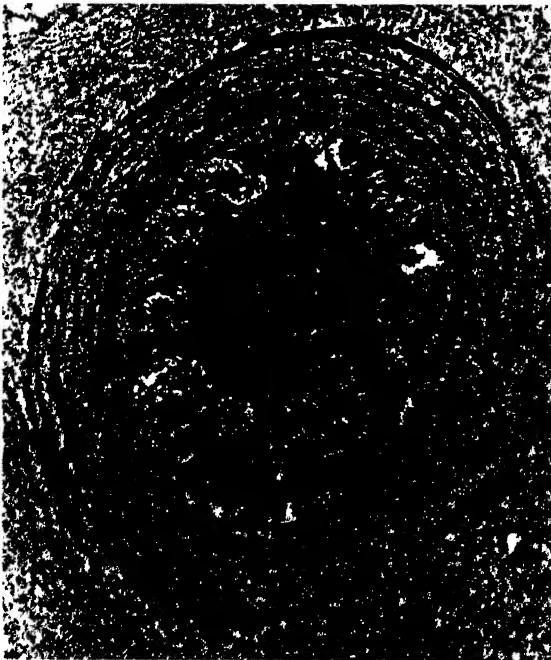


Fig. 14.26 Thrombophlebitis in Buerger's disease. *Left*, a superficial vein, showing inflammation of the wall, thrombosis and early organisation. Several multinucleated giant cells lie in and adjacent to the thrombus. $\times 60$. *Right*, an older lesion with more advanced organisation of the thrombus: note the giant cell (*below centre*) and pleomorphic inflammatory infiltrate. $\times 250$.

developing intermittently over a period of years. It may involve mainly either arteries or veins, but usually both. The changes of ischaemia, including gangrene of the extremities of the affected limbs, eventually result.

Aetiology. Features which distinguish Buerger's disease from atheroma with superadded thrombosis are its relatively early onset, inflammatory nature, predilection for smaller vessels, involvement of veins as well as arteries and of the upper limbs as well as the lower, and its rarity in women. Suggestions that the lesions are simply a variant of atheroma are almost certainly mistaken, and have probably arisen from the examination of limbs amputated after prolonged disease, when the lesions are burned-out and there is co-incidental atheroma. It is also a misconception that the disease occurs especially in Jews, but a high incidence of HLA antigens A9 and B5 in sufferers has been reported, suggesting a genetic predisposing factor.

The inflammatory nature of the early lesions suggests a specific causal agent (as was postulated by Buerger) but none has been detected. The single known important predisposing factor is cigarette smoking. The disease is practically confined to heavy smokers, and there is a strong clinical impression that its progress is arrested or diminished by giving up smoking. Hypersensitivity to tobacco proteins has been suggested as a causal factor.

Clinical features. The symptoms are varied and depend on the degree of arterial obstruction. The earliest are pain, paraesthesia and circulatory disturbances e.g. local redness which disappears on elevating the limbs. On walking there is often cramp-like pain and inability to progress—'*intermittent claudication*'; this is a result of ischaemia of the calf muscles and occurs in other forms of arterial disease. Later, more severe trophic changes appear, including intractable ulceration, and gangrene which is apt to spread slowly (Fig. 14.27); amputation, sometimes repeated, is often necessary but the need for surgery may be minimised by therapy which improves the collateral circulation. In view of the widespread involvement of the arteries, amputation, if required, should be performed at a high level. If the vessels of the arms are severely affected, the condition may simulate Raynaud's disease in the male.



Fig. 14.27 Gangrene of part of the foot and toes in a case of thromboangiitis obliterans.

Polyarteritis nodosa (*Periarteritis nodosa*)

The lesions of this disease consist of multiple foci of necrosis, inflammation and usually thrombosis, followed by healing, in the walls of medium sized and small arteries and arterioles. Vessels in any part of the body may be affected, and there is involvement of many organs and tissues. In its most severe form, with many simultaneous acute lesions, it is rapidly fatal from haemorrhage due to rupture of a weakened artery, or more often from ischaemia of vital organs. Most cases are, however, chronic, with acute lesions developing over years, and usually causing death from ischaemic effects. The condition occurs over a wide age range, but predominantly between 20 and 40.

Pathological findings. The early lesion consists of a focus of fibrinoid necrosis of the media and intima of a small or medium sized artery (up to about 3 mm diameter) or an arteriole. Necrosis is accompanied by acute inflammation with polymorph infiltration (often including eosinophils) of the whole thickness of the vessel wall and particularly intense in the adventitia and surrounding tissue (Fig. 14.28). Lesions affect the whole circumference of smaller arteries, but often only a segment of the wall of the larger vessels (Fig. 14.29). Occlusive thrombosis is common in the acute stage, but in some cases there is severe haemorrhage. The acute changes progress to more chronic inflammation, with replacement of the necrotic vessel wall by fibrous tissue infiltrated with lymphocytes, plasma cells and macrophages, and the thrombus undergoes organisation. The weakened wall may stretch to form an aneurysm (Fig. 14.30),



Fig. 14.28 An acute lesion of polyarteritis nodosa in a small artery in the kidney. There is fibrinoid necrosis and an intense inflammatory cellular infiltrate in and around the wall of the artery. $\times 100$.

but even without this the healed lesion may project as a nodular thickening of the vessel wall, and microscopy then shows a sharply-defined zone of fibrous replacement of the artery wall.

The lesions are multiple; they occur in almost any small or medium-sized artery or arteriole, but are commonest in those of the kidneys, heart, gut, liver, pancreas and nervous system, and in the skeletal muscles. Their effects, apart from haemorrhage, are due to acute or chronic ischaemia, and depend on the distribution of the lesions and arterial anastomoses in particular sites. Infarcts and patches of chronic ischaemic atrophy result in the heart, kidneys, etc.

The disease may be severe and progress rapidly to death, but more often the course extends over some years, with periods of quiescence alternating with the development of new lesions. In most cases, death eventually results from lesions in the kidneys, heart or other vital organs.

Clinical features depend on the number and sites of lesions, and as these are widespread and vary greatly in their distribution, the clinical features also show great variation. In severe cases there is fever, prostration, neutrophil (and sometimes eosinophil) leukocytosis and a very high ESR. In less acute cases the disease fluctuates, with quiescent periods and exacerbations. Lesions in the peripheral nerves (Fig.



Fig. 14.29 Polyarteritis nodosa involving a coronary artery. The lesion is less acute than in the previous figure. Part of the circumference of the vessel wall (*above*) has been severely damaged, with interruption of the internal elastic lamina (stained black) and replacement of the inner part of the wall by fibrous tissue. There is more diffuse inflammatory cellular infiltrate. $\times 70$. (The late Dr. Janet Niven.)



Fig. 14.30 Transverse section of the kidney of a patient who died of acute polyarteritis nodosa. In this instance, the necrotising lesions have resulted in aneurysmal dilatations, together with thrombosis. This has led to multiple infarcts. $\times 1.4$.

21.66, p. 782) result in paraesthesias, and symptoms may arise from ischaemia of virtually any tissue. Angina, cardiac failure, renal failure and hypertension are among the commoner clinical manifestations, but infarction of the gut, etc., can also cause death.

Diagnosis depends on suspecting the disease from the clinical features, blood changes, etc., and confirmatory biopsy. The choice of tissue for biopsy depends on the clinical features, but confirmation is often obtainable from skeletal muscle biopsy, particularly if tissue is removed from a tender spot in a muscle. Inflammatory changes are much more severe than in the necrotising arteriolar lesions of malignant hypertension.

Aetiology. The arterial lesions of experimentally-induced acute immune-complex disease (p. 154) resemble those of polyarteritis nodosa. To explain the development of acute lesions over a prolonged period, as in polyarteritis nodosa, it is necessary to assume that fluctuations in the plasma levels of postulated antigen and antibody results, from time to time, in the formation of heavy concentrations of immune complexes in the circulation. In support of the disease being due to a hypersensitivity type 3 reaction, immunoglobulins and products of activation of complement have been demonstrated in the acute lesions. The disease sometimes complicates systemic lupus erythematosus and the lesions in such cases are presumably due to complexes of auto-antibodies with DNA, etc. (p. 164). Approximately 50 per cent of cases arise in chronic carriers of hepatitis B virus with circulating HBsAg-Ab complexes, and HBsAg has been demonstrated in the acute lesions in some such cases. In many cases, however, the postulated antigen has not been identified, although there is often an association with various drugs, notably sulphonamides and penicillin.

Variants of polyarteritis nodosa

A **microangiopathic form** of polyarteritis nodosa affects mainly arterioles and very small arteries in the kidneys and elsewhere, and gives rise to microangiopathic haemolytic anaemia (p. 531) and uraemia, sometimes without hypertension: the renal lesions include **focal and rapidly progressive glomerulonephritis** (p. 821). In many instances the condition appears to be a hypersensitivity reaction to drugs or micro-organisms.

Wegener's granulomatosis is a variant in which lesions in vessels in the nasopharynx result in an ulcerating granulomatous lesion, and lesions also tend to occur especially in the lungs and kidneys.

Localised polyarteritis, with lesions indistinguishable from polyarteritis nodosa, occurs in the gall-bladder and appendix, and has an excellent prognosis. **Necrotising vasculitis of the skin** is a complex subject (p. 1065). The lesions of polyarteritis nodosa

can remain confined to the skin for some years, but there are other, localised forms of dermal vasculitis which are distinct from polyarteritis nodosa.

Other forms of arteritis

Idiopathic aortitis in Africans. An inflammatory lesion affecting all parts of the aorta has been described in young Africans. There is infiltration of the adventitia and media with lymphocytes and plasma cells, and destruction of the elastic tissue, ending in dense collagenous fibrosis. The mouths of the renal arteries are often involved, resulting in unilateral or bilateral renal artery stenosis often leading in turn to hypertension. Micro-organisms have not been demonstrated and the aetiology of the lesion is unknown.

Tuberculosis. Marked periarteritis and endarteritis sometimes occur in relation to tuberculous lesions. They are often a prominent feature in tuberculous meningitis, especially with late or inadequate treatment (Fig. 21.33, p. 751). These changes may lead to complete obstruction of the vessel and cerebral infarction. Occasionally, however, the wall of an artery adjacent to a pulmonary tuberculous cavity may be weakened before obliteration occurs and an aneurysm may form (p. 478).

Rheumatic arteritis. Lesions similar to those found in the heart in rheumatic fever occur in the walls of large arteries. In the aorta they commence in the adventitia and consist of an infiltration of the tissues with lymphocytes and plasma cells. There may be foci of histiocytes, and typical Aschoff bodies with characteristic cells (p. 412) may form. The cellular infiltration may spread into the media and lead to absorption of elastic tissue, but this rarely extends beyond the outer third of the media. Such lesions do not weaken the wall sufficiently to produce aneurysms.

In the smaller arteries, rheumatic lesions of various distribution have been found, especially in the visceral branches. They are acute and may be accompanied by necrosis of the media as well as by leukocyte infiltration; they thus resemble the lesions of polyarteritis nodosa but thrombosis and aneurysms have not been found. Further work is required to establish the relation of rheumatism to disease of the smaller arteries. The lesions are clinically silent, but they may form the basis of the subcutaneous lesions which commonly occur in rheumatic fever and illustrate its systemic nature.

Aortitis in ankylosing spondylitis. Unexplained lesions of the aortic valve and ascending aorta, identical in appearance to syphilitic arteritis but without serological evidence of syphilis, sometimes develop in males with ankylosing spondylitis.

Giant-cell or temporal arteritis. This is a fairly uncommon condition, occurring mostly in old people of both sexes. It affects mainly arteries of the head,

but is sometimes much more widespread, and the aortic arch and its major branches are occasionally involved. Diagnosis is often based on clinical examination and biopsy of the temporal artery, which is conveniently superficial and often affected.

The lesion is an inflammation of the whole thickness and whole circumference of the affected arteries, affecting either a continuous length of the vessel or appearing as multiple focal lesions along it. The vessel wall is infiltrated with leukocytes (mainly polymorphs) in the early stages, but the subsequent reaction is granulomatous, with accumulation of lymphocytes, macrophages and multinuclear cells (of both Langhans' and 'foreign-body' types) which sometimes appear to develop in relation to fragments of the disrupted internal elastic lamina. Fibrous thickening of the intima, fibrous replacement of the media, and commonly thrombosis and organisation, result in a severely scarred vessel with a narrowed or obliterated lumen (Fig. 14.31).



Fig. 14.31 Section of the temporal artery in giant-cell arteritis, showing multinucleated giant cells lying in relation to the internal elastic lamina (now disrupted and seen only as small fragments). There is gross intimal thickening, possibly from organisation of thrombus, and a very narrow lumen. $\times 120$.

Clinically there may be localised reddening of the skin over an affected vessel, which is usually tender or painful and sometimes nodular. Depending on which arteries are involved, there may be headache, visual disturbances and even blindness (from involvement of the retinal arteries), facial pain, and sometimes cerebral infarction and other features

resulting from more extensive arteritis. In some instances, the disease occurs in association with polymyalgia rheumatica. It is of entirely unknown aetiology and usually self-limiting.

Takayasu's disease. This is a rare condition, first reported from Japan, in which the aorta and the large arteries arising from the aortic arch are affected by an arteritis which may resemble syphilitic aortitis, caseating tuberculosis or temporal arteritis. Intimal thickening, sometimes with superadded thrombosis, severely narrows or occludes the subclavian, carotid and innominate arteries (hence the term *pulseless disease*), with resulting ischaemia of the head and arms. The disease affects mainly young women, and the aetiology is unknown. Occasional cases with vascular occlusions suggestive of Takayasu's arteriopathy are encountered in Britain, but the patients are usually older, and may be of either sex: these cases are attributable to severe atheroma or syphilitic arteritis of the major arteries.

Thrombotic microangiopathy (thrombotic thrombocytopenic purpura)

This is characterised by the deposition of homogeneous eosinophilic material, at least some of which is fibrin, in the intima and lumen of visceral arterioles, without an associated inflammatory reaction. It sometimes accompanies, and may belong to, the group of connective tissue diseases. Further details are given on p. 531.

Raynaud's disease

Nomenclature. In 1862 Maurice Raynaud described a series of cases of intermittent impairment of the circulation through the extremities, usually presenting as an abnormal response to exposure to cold. It has since become apparent that these effects can occur in subjects with or without organic vascular disease. There has been considerable confusion over nomenclature, but it is now customary to apply the term *Raynaud's disease* to cases apparently due wholly to abnormal angiospasm and to group together under the term *Raynaud's phenomenon* cases in which organic vascular changes play a major role.

Raynaud's disease occurs mainly in women, usually starting in adolescence and often continuing indefinitely. It usually affects the fingers, but occasionally the tip of the nose, ears

and toes. Involvement is symmetrical, e.g. the fingers in both hands are equally susceptible. On exposure to cold the fingers become cold, white or cyanotic and may be numb or extremely painful. The circulation is restored by warmth, but trophic changes may occur in the skin and whitlows are common: ulceration and gangrene seldom occur unless exposure to cold is prolonged.

Histological reports on the condition are few, because material is not usually excised unless gangrene develops. Thickening of the digital arteries has been reported, but these vessels are normally very thick walled, and reports of thickening have usually been erroneous. In cases where gangrene develops, however, there may be thrombosis and recanalised vessels may be found (Fig. 14.32).

The condition appears to be an exaggeration of the normal response to exposure to cold. Pre-ganglionic sympathectomy is not curative.

Raynaud's phenomenon. This consists of symptoms similar to those of Raynaud's disease, but attributable to organic vascular disease. It may result from a number of different conditions, including thromboangiitis obliterans, use of vibratory power tools (e.g. pneumatic drills), atherosclerosis, systemic lupus erythematosus, cold-antibody auto-immune haemolytic anaemia and cryoglobulinaemia.

Involvement is not always symmetrical, and sympathectomy does not usually improve the condition. Both sexes are affected and, depend-

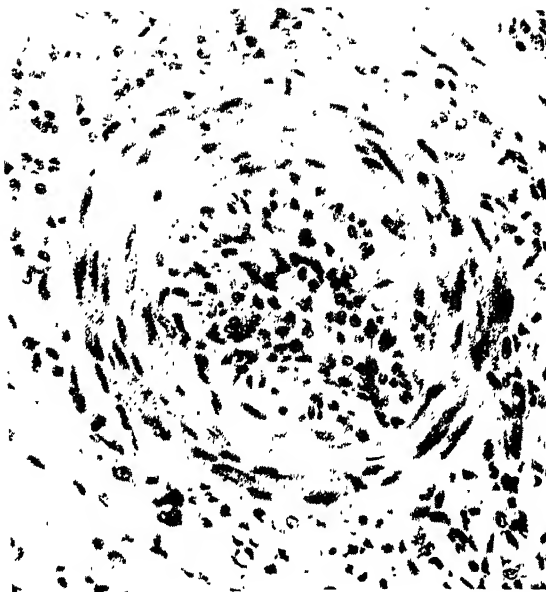


Fig. 14.32 Transverse section of a digital artery in Raynaud's disease. The lumen is obliterated by cellular fibrous tissue which has resulted from organisation of thrombus. $\times 250$.

ing on the causal condition, gangrene may develop.

Raynaud's phenomenon can result from vasoconstriction in response to cold superimposed on organic vascular disease, from arrest of the circulation in the extremities by cryoglobulins, and from agglutination of red cells in cold-antibody auto-immune haemolytic anaemia.

Aneurysms

Definition. An aneurysm is a local enlargement of the lumen of an artery.

Classification. A *true* aneurysm is formed by local dilatation of an artery, the blood being contained by the stretched vessel wall, or, when this is no longer recognisable, the surrounding connective tissue. A *diffuse* or *fusiform* aneurysm involves the whole circumference of the wall symmetrically whereas the *saccular* type is an asymmetrical bulge communicating with the artery through an aperture which does not in-

volve the whole circumference. These terms may not be applicable to advanced lesions, which are often very irregular in form.

The term *dissecting aneurysm* is used to describe a condition in which the wall of an artery (usually the aorta) splits, and blood tracks along the media, separating the inner from the outer layers. Some other lesions are described as aneurysms (p. 389).

Pathogenesis. The force which expands an aneurysm is the blood pressure, but for an an-

eurysm to form there must be an arterial lesion which weakens the media locally. Stretching usually results in further weakening, so that once an aneurysm has started it tends to expand and commonly ruptures. Occasionally thrombus forms in thick layers which fill the whole sac (Fig. 9.19, p. 239).

Atheromatous aneurysm

In Europe and N. America, atheroma is now the most common cause of true aortic aneurysm, due to the decline and early treatment of syphilis and the concurrent increase in atheroma. Atheromatous aneurysms occur usually after the age of 50, and much more commonly in men than women. They are usually fusiform (Fig. 14.33) and may rupture while still quite small. The aneurysm forms as a result of pressure atrophy of the media over atheromatous



Fig. 14.33 Atheromatous aneurysm of the abdominal aorta arising below the origins of the renal arteries. The aneurysm, which is fusiform, has been repaired by a dacron tube, but death resulted from haemorrhage from rupture at the suture line.

plaques or actual extension of the plaque into the media. The microscopic changes seen at the edges of the aneurysm are those of atheroma, sometimes with a marked leukocytic reaction around the fatty debris, and there may be some lymphocytic infiltration round the vasa vasorum in the adventitia and media. These aneurysms are usually a complication of severe atheroma and affect especially the abdominal aorta or a common iliac artery. They usually arise below the level of the renal arteries.

Effects. Large aneurysms are very liable to rupture with retroperitoneal haemorrhage, the clinical features being those of an acute surgical abdominal emergency. In some cases thrombosis of the aneurysmal sac results in ischaemia in the legs, kidneys, etc. Pressure effects are not conspicuous.

Syphilitic aneurysm

This occurs as a complication of syphilitic aortitis, usually above the age of 40. Large aortic aneurysms were previously due in most instances to syphilis, but are now very rare as a result of successful treatment in the primary and secondary stages of the disease. The commonest site is the aortic arch, because it is the part most frequently affected by syphilitic mesoaortitis (p. 377), of which aneurysm is a complication. The focal loss of elastica and muscle in the media results in weakening of the wall, and there may be diffuse dilatation of the ascending aorta and arch: more localised stretching results in a fusiform or saccular aneurysm, which is often accompanied by smaller aneurysmal bulgings, along with the stellate scars and intimal thickening characteristic of syphilitic mesaortitis. As an aneurysm forms, the elastic tissue and muscle of the artery wall soon degenerate and the sac comes to be composed of layers of fibrous tissue, on which laminated thrombus forms (Fig. 9.19, p. 239). Blood may infiltrate the wall of the aneurysm and ooze for some distance into the tissues around; accordingly the limits of the aneurysm are badly defined.

Effects. Pressure on surrounding structures leads to the syndrome of superior mediastinal compression; the great veins may be displaced and undergo thrombosis, resulting in congestion of the head and neck and enlargement of

collateral veins. Involvement of the oesophagus may cause dysphagia, while pressure on a major bronchus may cause a chronic cough and suppurating bronchopneumonia. Aneurysms of the transverse part of the aortic arch may compress and stretch the left recurrent laryngeal nerve and cause paralysis of the left vocal cord. Rigid structures such as the bodies of vertebrae may be eroded and the bare bone come to form part of the wall of the sac; the intervertebral discs offer greater resistance to absorption and persist longer.

Rupture of an aneurysm may occur into practically any tube or cavity in its neighbourhood, and occasionally takes place externally through the chest wall.

Embolism from thrombus formed within an aneurysm is uncommon.

Cardiac hypertrophy and dilatation occur only when the syphilitic mesaortitis results in aortic-valve incompetence. Otherwise aortic aneurysms, even very large ones, do not affect the heart as there is no interference with cardiac output.

Dissecting aneurysm

This is now the commonest cause of rupture of the aorta. In most cases it results from rupture of the inner part of the wall of the aorta (Fig. 14.35) often due to degenerative changes in the media in which the elastica and muscle are replaced, in an irregular, patchy manner, by a metachromatic mucoid substance (Fig. 14.34) and the surviving elastica often appears fragmented. These changes are known as **Erdheim's medial degeneration**. Sometimes there are small areas of necrosis with softening (*medionecrosis*) but in our experience this is rare. The cause of these changes is unknown and in some cases of dissecting aneurysm they are not apparent. Dissecting aneurysm is commonest in people over 40 years old, usually with high blood pressure, but it occurs also in young people without hypertension, and is commoner in men than in women.

The event which causes dissecting aneurysm is a sudden transverse tearing of the inner part of the aortic wall, through which blood from



Fig. 14.34 Medial degeneration of the aorta. There are gaps in the elastic tissue of the media, which is stained black in the left illustration. In the right photomicrograph, myxoid ground substance is stained black and is patchily increased. The patient was a woman of 24 who died of spontaneous rupture of the aorta. $\times 50$.

the lumen penetrates into the weakened media. The tear is usually in the ascending aorta and may extend around almost the whole circumference (Fig. 14.35). Much less commonly, it occurs just distal to the point of insertion of the ductus arteriosus. It tends to occur during physical exertion and is probably brought about by the frictional tractive force of the blood during systole, which thrusts the intima in the same direction as the blood flow. Because of the weakened media, the inner part of the wall slides to and fro with each heart-beat and eventually ruptures. Blood tracks between the inner two-thirds and outer third of the media, effectively dissecting the inner part of the wall from the outer part.

In some cases, the outer part of the aortic wall is ruptured almost immediately by the pressure of blood entering the media, with fatal haemorrhage. Occasionally the dissection re-



Fig. 14.35 Dissecting aneurysm of the aorta. A rupture of the inner part of the wall, above the aortic valve, extends almost around the circumference, and blood has tracked up between the inner and outer parts of the wall, producing a space, *a*, the so-called dissecting aneurysm. $\times 1$.



Fig. 14.36 Localised dissecting aneurysm; the space in the media is filled with dense thrombus. $\times 1.2$.

mains localised (Fig. 14.36) and clotting and organisation of the blood in the media results in healing. In most cases, however, blood tracks proximally and distally in the media. Commonly it reaches the aortic ring and ruptures into the pericardial sac, causing death from cardiac tamponade. It may also track distally as far as the abdominal aorta and even into the iliac arteries, and rupture may result in haemorrhage into the mediastinum, pleura, retroperitoneal tissue or peritoneal cavity.

When dissecting aneurysm extends to the origins of branches of the aorta, it compresses and narrows the lumen of the branches and may extend along them: this can result in obstruction of the coronary arteries, branches of the aortic arch, mesenteric and renal arteries, etc., obviously with serious effects.

Occasionally a second tear develops, usually in the abdominal aorta, through which blood in the dissecting aneurysm re-enters the lumen. If the patient survives, the channel in the media may become lined by endothelium supported by fibrous tissue and 'double-barrelled' aorta results (Fig. 14.37).

Dissecting aneurysm is generally accompanied by severe tearing pain in the chest. If untreated, it is usually fatal within a few days, and even with surgical treatment the outlook is poor.

Marfan's syndrome. This is a disorder of connective tissue, inherited usually as an autosomal domin-

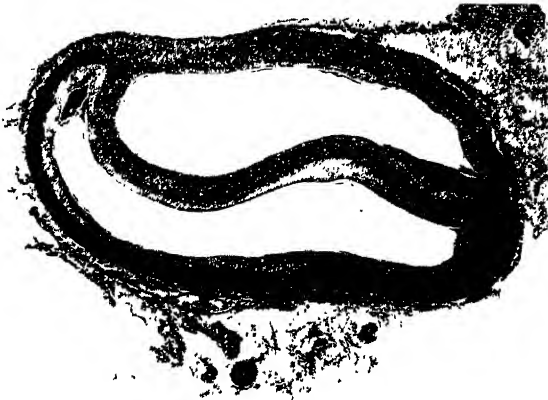


Fig. 14.37 Dissecting aneurysm of the aorta. In this case the blood burst back into the lumen through a second rupture, giving a 'double-barrelled' aorta, seen here in cross section. $\times 2.5$.

ant, and characterised by laxity of ligaments, e.g. of joints and of the eye lens, inadequate elastic fibre formation in the aorta, tall slim build, long tapering fingers (*arachnodactyly*) and other skeletal abnormalities. Disturbances of vision result from subluxation of the lens, and there may also be deafness. The aortic media lacks elastic fibres, the appearances resembling those in Erdheim's medial degeneration. Fusiform aneurysm or dissecting aneurysm may result, and at necropsy there may be multiple healed aortic intimal tears.

Other causes of aortic rupture. Apart from aneurysm, rupture of the aorta may result from damage to its wall from outside, as by the perforation of an impacted fishbone or other sharp foreign body in the oesophagus; also from very severe injury such as crushing of the chest. In children traumatic rupture can occur without fracture of the ribs. Carcinoma of bronchus or oesophagus may invade the aortic wall and cause fatal haemorrhage.

Infective (mycotic) aneurysm

This may occur at the beginning of the aorta by direct extension of micro-organisms from vegetations in bacterial endocarditis, particularly *Staphylococcus aureus* (Fig. 15.32, p. 423). The organisms settle on the intima, an infective thrombus forms, invasion and weakening of the wall follow, and an *acute aneurysm* is produced, which may rupture; occasionally multiple aneurysms are present. In smaller arteries infective aneurysms result from lodgment of small infected emboli in the vasa vasorum, rather than from the presence of an infected embolus in the lumen of the artery. Inflammatory softening of the arterial wall results. Infective aneurysms may occur in a limb or viscus, and the effects are similar to those seen in the non-infective

aneurysms of polyarteritis nodosa (p. 380). Infected emboli in the lumen of an artery may give rise to acute inflammatory softening with rupture and cerebral haemorrhage, e.g. in staphylococcal pyaemia.

A mycotic aneurysm is sometimes seen in the wall of a tuberculous pulmonary cavity and may cause fatal haemoptysis. Usually, however, occlusion of the vessel by endarteritis obliterans or thrombosis prevents aneurysm formation.

Cerebral aneurysms

Berry aneurysms of the circle of Willis and its branches occur at all ages and appear to be due mainly to congenital weakness of the arterial wall. Evidence of this is destroyed when the aneurysm forms, but a deficiency in the medial muscle, especially in the acute angle between large branches, is demonstrable in other arteries at the base of the brain, and similar deficiencies are found in a small proportion of people without aneurysms. Probably as a result of stretching, the internal elastic lamina is absent in much of the wall of the aneurysm, which consists of a thin, often transparent layer of hypocellular fibrous tissue. Although often called *congenital*, most of these aneurysms develop from adolescence onwards. The muscle defect, which is congenital, and hypertension are contributory causes of aneurysm formation and rupture in older people. The aneurysms usually occur singly, but may be multiple: they are usually less than 1 cm diameter, but may be much larger. The commonest site is at the origin of the middle cerebral, followed by the anterior communicating artery (Fig. 14.38), but they may arise practically anywhere on the circle of Willis and its major branches. Rupture of these aneurysms is the principal cause of spontaneous subarachnoid haemorrhage, but in some cases the sac is buried in the cortex and they bleed into the brain. The effects of rupture are described more fully on p. 740. Quite commonly, an aneurysm becomes partly or completely occluded by thrombus (Fig. 14.39). Occasionally a fusiform aneurysm of the basilar artery develops as a result of atheroma.

Micro-aneurysms. The occurrence of multiple micro-aneurysms on the small cerebral arterial twigs in hypertensive subjects was described over a century ago, but they are difficult to find without special techniques, and only recently has their common occurrence been reaffirmed. In a painstaking study, Cole and Yates (1967)



Fig. 14.38 Base of the brain showing subarachnoid haemorrhage which resulted from rupture of a berry aneurysm of the basilar artery (not shown). A second, intact berry aneurysm (*black arrow*) is seen on the anterior communicating artery. (The anterior cerebral arteries are marked by white arrows.)

using a micro-angiographic necropsy technique reported the occurrence of multiple (usually 15–25) aneurysms of up to 2 mm diameter, occurring mainly on arteries of less than 250 μ m diameter. They were detected in over 50 per cent of hypertensives over 50 years of age, and the incidence increased with age: in normotensives, the incidence was low, aneurysms being found only in a few subjects over 65 years old. The aneurysms were most numerous in and around the basal ganglia, and occurred usually at or near branchings of the striate arteries: they were found also in the subcortical white matter and in the mid-brain and cerebellum.

The aneurysms may be saccular or fusiform, and the adjacent artery and wall of the sac show hyaline thickening of the intima, sometimes with fibrinoid change: the internal elastic lamina is usually absent from, or fragmented in, the wall of the sac, and muscle is usually absent. Thrombus, sometimes organised, may fill the sac, and there is often evidence of old or



Fig. 14.39 Aneurysm of circle of Willis, almost completely filled with thrombus. $\times 7.5$.

recent leakage of blood into the surrounding tissues.

Cole and Yates detected micro-aneurysms in 18 of 20 hypertensives dying from cerebral haemorrhage, and they provide evidence which suggests strongly that rupture of such aneurysms is the usual cause of cerebral haemorrhage in hypertensive subjects (p. 740).

Other forms of aneurysm

Injury to the wall of an artery by a stab wound, etc., may result in the development of a **traumatic aneurysm**. It is difficult to understand how this can result from an injury which penetrates the whole thickness of the wall, and it seems more likely to be due to stretching of the fibrous scar resulting from healing of an incomplete laceration of the vessel wall.

Injury of an adjacent artery and vein may result in an arterio-venous fistula: in some instances the connection is by a channel with a fibrous wall, which may dilate to form an **arterio-venous aneurysm**.

A **cirroid or racemose aneurysm** is a form of arterio-venous fistula which appears as a pulsatile swelling consisting of tortuous and dilated arteries and

veins with multiple intercommunications. The commonest site is the scalp, and it may cause pressure atrophy of the underlying bone. The condition is sometimes congenital, but more often the result of a blow on the head; some of the allegedly congenital

cases are probably the result of birth injury. Similarly a carotid-cavernous sinus aneurysm resulting from fracture of the skull base gives rise to great engorgement of the orbital veins and oedema of the orbit and conjunctiva.

Diseases of Veins

Compensatory enlargement of the veins takes place, as in the arterial system, when there is a sustained increase in blood flow, as in the uterine veins during pregnancy and in the collateral veins following obstruction of a major vein. As in the case of arteries, dilatation is followed by hypertrophy of the various elements in the wall of the vessels. Veins are, of course, not exposed to the marked variations of blood pressure which occur in arteries, but when they are subject to persistent over-distension, compensatory changes occur in their walls. There is little or no hyperplasia of the muscle, but the elastic tissue increases greatly and then undergoes degeneration; the fibrous tissue also increases and becomes hyaline (Fig. 14.40). Localised patches of thickening in the intima of veins are quite common and probably result from organisation of thrombus.

Acute thrombophlebitis

A distinction is sometimes made between *thrombophlebitis*, by which is meant a primary inflammatory condition of the vein, with secondary thrombosis, and *phlebothrombosis* (p. 239), in which a bland thrombosis of the vein occurs with, at most, mild preceding inflammatory change. In many instances the distinction is more theoretical than practical because the presence of thrombus in the lumen of the vein sets up reactive changes so that differentiation between mild thrombophlebitis and phlebothrombosis may no longer be possible. In *thromboangiitis obliterans*, however, veins are often affected and exhibit the same rather characteristic florid inflammatory lesions seen in arteries in this condition (Fig. 14.26, p. 379). The available evidence suggests that the inflammation is primary and the thrombosis a consequence of it.

Multiple venous thrombosis also occurs in

thrombophlebitis migrans but without arterial involvement: it usually affects superficial veins, but sometimes also deeper ones, in any part of

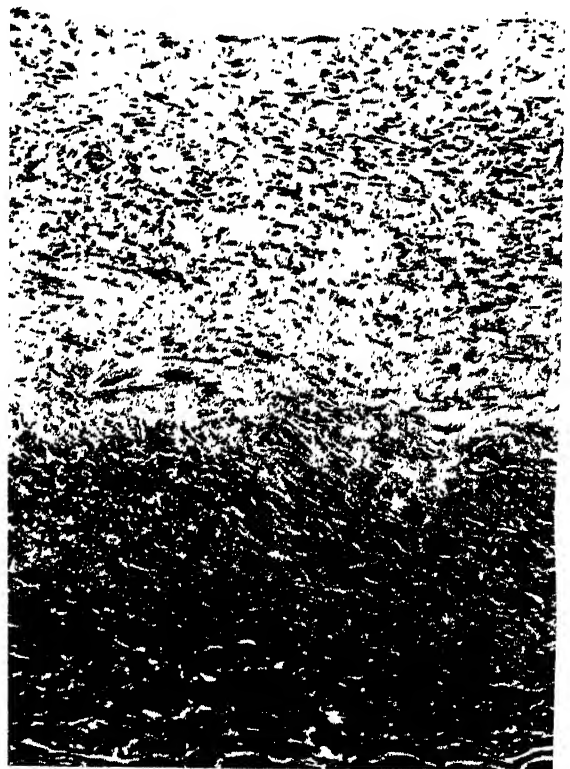


Fig. 14.40 Longitudinal section of the wall of a saphenous vein used to replace a length of an atherosclerotic coronary artery two years before death. The lumen is at the top and the dark fibres at the bottom are dense collagen fibres of the adventitia. The muscle coat of the media has undergone atrophy; the residual fibres are seen as an ill-defined dark layer near the periphery of the wall. Above this, the relatively acellular layer represents fibrosed and thickened media and the more cellular (inner) layer consists of greatly thickened intima which contains elastic fibres, collagen and large spindle cells. $\times 120$.

the body. In many cases the cause is not apparent, but in others it is associated with a carcinoma, most often of the pancreas but also of breast, stomach, bronchus, ovary, etc. The thrombotic episodes may be the first clinical manifestation of the cancer. In some cases, small vegetations form on the cardiac valves.

Outbreaks of a condition known as **tropical thrombophlebitis** have been reported in Africans: thrombosis is widespread and death may result from involvement of visceral veins. The cause is unknown.

Infective thrombophlebitis used to be seen commonly in the veins of the diploë and dural sinuses in middle-ear disease, in the uterine veins in puerperal sepsis, in the veins of the bone marrow in suppurative osteomyelitis, and occasionally in the pulmonary veins in cases of bronchiectasis. Veins involved in such lesions undergo thrombosis and the thrombus becomes invaded by bacteria, and then by polymorphs: fragments may break away and produce pyaemia.

In the rare condition known as *pylephlebitis suppurativa* (*portal pylephlebitis*) infection, e.g. from appendicitis, involves a small tributary of the portal vein and leads to progressive ascending thrombosis and suppuration, from which multiple abscesses in the liver may result.

Thrombophlebitis may also appear as a complication of conditions in which there is a bacteraemia, notably typhoid fever, and it is presumed that organisms circulating in the blood settle in the intima and produce an acute endophlebitis with secondary thrombosis.

Chronic phlebitis

Chronic inflammatory processes may spread to the walls of the veins and lead to reactive thickening; in fact, the smallest veins are affected in this way in all chronic inflammatory conditions. Chronic phlebitis of obscure origin is occasionally observed in the large vessels, for example in the portal vein, and may lead to thrombosis; cavernous tissue develops in the portal fissure and after some time it may be impossible to say whether the changes are congenital or secondary to the thrombosis.

Endophlebitis of the hepatic veins is the basis of the veno-occlusive disease of Jamaica and certain other tropical regions (see p. 664) and involvement of the hepatic ostia with throm-

bosis gives rise to the Budd–Chiari syndrome (p. 664). Hepatic endophlebitis also results from infestation with liver flukes and schistosomes.

Tuberculous invasion of veins most commonly results when a caseous lesion—usually in a pulmonary hilar lymph node—involves and destroys the wall of a vein: huge numbers of tubercle bacilli then enter the bloodstream and cause generalised miliary tuberculosis.

Veins are often invaded by **malignant tumours** which may then release cells singly or in groups, with the danger of metastatic growth in the lungs, etc. Cancer may also grow along the lumen of veins, an example being clear-cell carcinoma of the kidney, which commonly extends along the renal vein and even the inferior vena cava. Such invasion is usually accompanied by thrombosis.

Varicose veins

Dilatation and tortuosity of veins is termed *varicosity*. The changes may affect a group of veins diffusely or take the form of saccular dilatations. Varicosity of veins arises from chronic continuous or recurrent increase in the pressure of the blood within them, and this results mainly either from (a) the effects of gravity, e.g. in the leg veins, sometimes aggravated by compression proximally, or (b) obstruction of a major vein, leading to increased pressure in collateral veins.

‘**Gravitational**’ varicosity occurs in the saphenous system of the legs, notably the long saphenous vein. The condition is much commoner in women and there is a distinct *hereditary predisposition*. *Prolonged standing* upright without much muscular movement causes marked rise in pressure and distension of the long saphenous vein, for the valves can only play their part in breaking the venous pressure gradient between the heart and the foot if assisted by the pumping action of muscular activity of the lower limbs. Eventually the veins become permanently stretched, so that the valves are now incompetent, and even muscular activity does not protect the veins from increased pressure in the upright position; in consequence, stretching tends to progress and the veins become visibly swollen and tortuous, i.e. varicose. Venous stasis occurs in the legs due to the pressure of the gravid uterus on the iliac veins and without doubt *pregnancy* is a predisposing cause of varicose veins; this probably accounts for the

higher incidence in women, although *obesity* is also a predisposing factor. The venous valves and the muscle and elastic tissue of the vein walls atrophy somewhat irregularly so that thinning of the wall and pouchlike dilatations occur; finally the wall comes to be composed chiefly of fibrous tissue. The nutrition of the skin over varicose veins of the legs may be impaired. The skin becomes eczematous and pigmented, and chronic indolent '*varicose ulceration*' often follows: dilated veins involved in such ulcers may bleed severely, but this is easily stopped by raising the leg with the patient lying flat. Thrombosis is also apt to follow. Organisation of the thrombus is generally imperfect and it may become calcified.

Varicocele is another common example of 'gravitational' varicosity, in the pampiniform plexus of veins around the spermatic cord: it is commoner on the left side than on the right and various ingenious explanations have been suggested for this. The distended veins feel like a bag of worms. It often depresses spermatogenesis and impairs fertility, particularly if bilateral. This may be a temperature effect.

Haemorrhoids consist of varicosities of the haemorrhoidal venous plexuses, projecting from the surface just above or below the ano-rectal junction. They are common in preg-

nancy, probably due to the pressure of the uterus on pelvic veins, and also in people over 40 in whom constipation and straining at stool are causal factors. Portal hypertension, e.g. in cirrhosis, is also believed to be a predisposing factor.

Haemorrhoids may bleed and cause iron deficiency, or they may rupture into the perianal subcutaneous tissue, causing painful swellings. They may also become thrombosed or prolapse through the anal sphincter and become strangulated.

'Obstructive' varicosity. This is exemplified by chronic *obstruction to the portal venous blood flow* (due most commonly to cirrhosis or schistosomiasis of the liver), in which the vessels which form anastomoses between the portal and systemic venous systems become varicose (p. 692): the most important ones are those running longitudinally in the oesophageal and gastric submucosa (Fig. 19.22, p. 604), for they may rupture and bleed profusely. *Obstruction of the inferior vena cava* brings about dilatation of the veins of the abdominal wall, establishing a collateral circulation through the upper thoracic veins (Fig. 9.7, p. 231). *Obstruction of the superior vena cava* may occur in cases of bronchial carcinoma and leads to severe dusky cyanosis of the head, neck and arms, sometimes accompanied by pitting oedema of the hands.

Diseases of Lymphatic Vessels

The lymphatic vessels form a closed system separated by an endothelial layer from the tissue spaces. The walls of the small lymphatics are, however, extremely delicate, consisting mainly of a very thin endothelium and an incomplete basement membrane. Moreover, the junctions between endothelial cells are readily disrupted (p. 62). In consequence organisms, leukocytes and tumour cells readily pass into the lymphatic vessels; also red cells which escape from the capillaries by diapedesis may be present in large numbers in the lymphatics draining an inflamed area. The lymphatic vessels thus afford an easy means of communication between the tissues and lymph nodes. Involvement of the lymph nodes in this way occurs in two main conditions, **infections** and **tumours**, especially carcinoma. In both, the

extension may be due to transport of the organism or tumour cell by the lymph stream, i.e. metastasis in the strict sense. There may also be progressive involvement of the lymphatic vessels by the disease. In infections, this may involve either acute or chronic lymphangitis; in tumours, lymphatic permeation may occur, columns of cancer cells extending along the lymphatics (Figs. 12.21, 12.22, p. 333).

Acute lymphangitis. This is seen in pyogenic infections, and is a feature of erysipelas and infections of the hand, etc., due to haemolytic streptococci. The spread of infection along the lymphatics is sometimes accompanied by visible reddening of the overlying skin, with pain and tenderness and often swelling. Spreading lymphangitis is an important feature in puerperal sepsis and septic abortion and may

be followed by cellulitis of the loose connective tissue around the uterus. In other cases of bacterial infection, the organisms are carried by the lymphatic vessels and reach the lymph nodes without causing lymphangitis. A similar striking example is seen in bubonic plague, where even at the site of infection there is usually no inflammatory reaction, the first lesion appearing in the related lymph nodes.

Chronic lymphangitis occurs in various conditions; it may follow *repeated acute attacks of erysipelas*, and is an important feature in many types of *chronic inflammation*. In various chronic infections the spread of organisms by the lymphatics is of great importance. In *tuberculosis*, a disease which in the early stages may be regarded as essentially one of the lymphatic system, the organisms may be carried to lymph nodes without causing lesions on their way. They may, however, settle in the walls of the lymphatic vessels and give rise to tubercles which thus come to form rows along the vessels. In tuberculous ulceration of the intestine, small tubercles may be found along the lymphatics passing from the floor of the ulcer (Fig. 19.60, p. 633), and also in the mesenteric lymphatics. The thoracic duct may become involved by spread of bacilli along the lymph stream and ulceration of these lesions may set free a large number of tubercle bacilli into the circulation to set up acute miliary tuberculosis (p. 212).

In *syphilis* also, chronic lymphangitis is a prominent feature in connection with the primary lesion, and induration spreading along the lymphatics leads to the characteristic bubo in the regional lymph nodes.

Lymphatic obstruction: lymphoedema. Chronic obstruction of lymphatics may give rise to interstitial accumulation of lymph (lymphoedema). When this is prolonged there is proliferation of connective tissue in the lymphoedematous area, resulting in a firm, non-pitting oedema. The most striking examples are seen in *filariasis*, in which obstruction of major lymphatics, together with recurrent inflammation in the affected region, may lead to gross thickening of the tissues known as **elephantiasis**; the lower limbs and sometimes the male external genitalia may be involved (see below). In non-tropical countries, extensive carcinomatous permeation of lymphatics is a more common cause, but surgical removal of lym-

phatics or destruction of lymphatics by radiotherapy also cause lymphoedema. This is sometimes seen following radical mastectomy and radiotherapy of the axilla for breast cancer, where the arm may be sufficiently deprived of its lymphatic drainage to develop gross lymphoedema without carcinomatous involvement of lymphatics. Several instances of tumour growth resembling lymphangiosarcoma have been observed in the lymphoedematous arm. It is not clear whether they are true sarcomas, or metastatic breast carcinoma the appearance of which is modified by the lymphoedematous environment.

Filariasis

A number of species of Filarioidae, a family of nematode worms, infest man in tropical and sub-tropical countries. Infestation results from transmission of larvae in the bite of infested mosquitoes. The larvae mature into adult worms in the human host, and the female produces *microfilariae* which are transmitted to man-biting female mosquitoes, in which the life-cycle is completed.

Bancroftian filariasis, caused by *Wuchereria bancrofti* is the most widespread and important form of filariasis. It occurs in West, Central and East Africa, Egypt, parts of Central and South America, and in South East Asia.

The adult worms are filiform, white and show wriggling movements: the female is almost 80 mm long, 0.3 mm wide, and the male is smaller and thinner with a spirally twisted tail. The adults colonise the lymphatic vessels and the sinuses of lymph nodes, particularly in the inguinal region, spermatic cord and upper arm: they may also infect the para-aortic lymph nodes and sometimes the thoracic duct.

Acute inflammatory reactions, possibly due to hypersensitivity, may develop in relation to the adult worms, involving the spermatic cord or epididymis, testis and inguinal lymph nodes. A more severe granulomatous reaction occurs when the adult worms die and disintegrate: granulation tissue containing tubercle-like follicles forms around the dead worms and extends to involve adjacent tissues and blood vessels, often with resultant venous thrombosis. This chronic inflammatory reaction results in obliteration of the lymphatics with consequent lymphoedema and permanent and sometimes en-

ormous swelling (**elephantiasis**) which may affect the lower limb, scrotum, vulva and occasionally the arm or breast. Chronic hydrocele may also develop. The lymphatics in the lymphoedematous tissue are distended and the skin is at first tense and shiny, later scaly and rough. Secondary bacterial infection and ulceration may occur. If the thoracic duct or abdominal lymph nodes are involved, there may be chylous ascites (escape of abdominal lymph into the peritoneal cavity).

The diagnosis of infestation may be made by detecting microfilariae in the peripheral blood, bearing in mind their presence in the bloodstream mainly at night. The microfilariae are unsheathed, about 250 mm long, and are rendered conspicuous in a wet film of peripheral blood by the commotion of adjacent red cells caused by their vigorous movements. Methods for concentrating microfilariae in the blood are available. Certain differentiation from other microfilariae requires examination of a stained preparation. Microfilariae may also be detected by skin biopsy (Fig. 14.41). In chronic cases with elephantiasis, there may be no surviving adult worms and microfilariae often cannot be found. Serological and skin tests are also used in the diagnosis.

A somewhat similar form of filariasis is caused in Asia by *Brugia malayi*, in which the distribution of the adult worms and so of the lesions is different.

Loiasis, due to the filaria *Loa loa*, occurs in West and Central Africa. The adult worms live and move around in the connective tissues throughout the body and are occasionally seen wriggling in the conjunctiva! They induce superficial inflammatory lesions (fugitive or Calabar swellings) which persist for a few days and are probably due to a hypersensitivity reaction. Heavy infestation is associated with fever, vague pains, ill-health and sometimes



Fig. 14.41 Microfilariae of *Wuchereria bancrofti* in a skin biopsy. $\times 200$.

intense itching. The unsheathed microfilariae appear in the blood during the daytime, presumably as an adaptation to the *Chrysops* vector mosquito which hunts man in daylight.

Onchocerciasis is caused by infection with *Onchocerca volvulus* and occurs in equatorial Africa and tropical America. It is transmitted to man by *Simulium* flies which live along the banks of rivers. The adult worms are entwined in groups often ensheathed in fibrous tissue, causing subcutaneous swellings, particularly over bony projections in the lower limbs and, in tropical America, in the scalp. The female is about 500 mm long and the male is much smaller.

Microfilariae are produced from about a year after infestation, and migrate in the dermis and subcutaneous tissue. They do not circulate in the blood, but have a predilection for the conjunctiva, the chambers of the eye and optic nerve, where their presence may result in blindness.

References

- Benditt, E. P. (1977). Implications of the monoclonal character of human atherosclerosis plaques. *American Journal of Pathology* **86**, 693–702.
- Cole, F. M. and Yates, P. O. (1967). The occurrence and significance of intracerebral micro-aneurysms. *Journal of Pathology and Bacteriology* **93**, 393–411.
- Goldstein, J. L. and Brown, M. S. (1977). The low-density lipoprotein pathway and its relation to atherosclerosis. *Annual Review of Biochemistry* **46**, 897–930.
- Kannel, W. B. (1971). Serum lipid precursors of coronary heart disease. *Human Pathology* **2**, 129–51.
- Kannel, W. B., McGee, D. and Gordon, T. (1976). A general cardiovascular risk profile. *American Journal of Cardiology* **38**, 46–51.
- Smith, J. P. (1956). Hyaline arteriosclerosis in spleen, pancreas and other viscera. *Journal of Pathology and Bacteriology* **72**, 643–56.
- World Health Organisation (1970). Classification of hyperlipidaemias and hyperlipoproteinaemias. *Bulletin of the World Health Organisation* **43**, 891–908.

Further Reading

- Atherosclerosis—a new look at the problem (1977). Symposium in the *American Journal of Pathology* **86**, 655–702.
- Ross, R. and Glomsett, J. A. (1977). The pathogenesis of atherosclerosis—a new look. *New England Journal of Medicine* **297**, pp. 369–77, 420–25.
- Woolf, N. (1978). The pathogenesis of atherosclerosis. In *Recent Advances in Histopathology*, No. 10, pp. 45–67. Edited by P. P. Antony and N. Woolf. Churchill Livingstone, Edinburgh, London and New York.

See also Further Reading for Chapter 15, p. 430.

The Heart

Disease of the heart now causes more deaths in Western countries than disease of any other organ, amounting to more than a third of all deaths in England and Wales according to the Registrar General's Statistical Review for 1977. Atheroma and thrombosis of the coronary arteries account for most of these deaths by causing ischaemic heart disease in its various forms. The other frequent causes of cardiac disease include systemic arterial hypertension, chronic diseases of the lungs which cause pulmonary hypertension, rheumatic and other lesions of the heart valves, congenital abnormalities, thyrotoxicosis and anaemia. The relative importance of these conditions varies greatly throughout the world.

The work of the heart. Assuming that at rest the stroke volume of the heart is 66 ml and the rate 72 beats per minute, the left ventricle has a minute volume of about 5 litres, and a daily

output of 7200 litres (about $7\frac{1}{2}$ tons). The normal heart has great reserve power, and this can be substantially increased by physical training. During exertion, there is a greater venous return to the heart with consequent increase in diastolic filling and stretching of the muscle fibres; the response is a more vigorous contraction (Starling's law) and the blood pressure is raised. The rate of contraction also increases during exertion and these two factors together can raise the minute volume to about seven times that of the resting state.

This physiological performance can be maintained only if (1) the myocardium is intrinsically healthy, (2) the valves function efficiently and (3) the conducting system of the heart co-ordinates contraction of the chambers. Disturbance of any of these requirements can cause cardiac failure.

Cardiac Failure

Definition. Cardiac failure is that state in which the ventricular myocardium fails to maintain a circulation adequate for the needs of the body despite adequate venous filling pressure. Failure of one or both atria is common, but the effect on cardiac function is relatively unimportant unless ventricular function is also deficient.

Causes

Cardiac failure is due to weakness or inefficiency of myocardial contraction, to an abnormal increase of the work required of the myocardium, or to a combination of both. These two basic causes may be further classified as follows.

(1) Intrinsic pump failure. This is most commonly due to *weakness of the ventricular contraction*, the commonest cause of which is myocardial ischaemia resulting from coronary artery disease. Other causes of myocardial weakness include myocarditis, severe toxic bacterial infections and congestive cardiomyopathy.

Systolic emptying of the affected ventricle(s) is incomplete because the force of contraction is reduced, and during diastole the chamber dilates to contain both the residual blood and that received from the atrium. The dilated chamber works at a disadvantage because the force required to provide a given pressure is greater in a large than in a small chamber. Consequently,

unless the cause is reversible, dilatation and failure tend to be progressive. Moreover, left or right ventricular dilatation results in stretching and incompetence (functional incompetence) of the mitral or tricuspid valve respectively, and this, as described below, increases the work of the dilated ventricle.

A less common cause of intrinsic pump failure is *impaired compliance of the myocardium* which in plain language means that the ventricles are too stiff to relax and fill properly during diastole, as in hypertrophic cardiomyopathy and amyloid disease of the heart. The abnormal rigidity may also interfere with myocardial contraction. By restricting cardiac filling, pericardial haemorrhage or effusion and restrictive pericarditis can produce similar effects.

Disorders of cardiac rhythm, resulting from various conditions, are also included in this group. Although minor irregularities such as sinus arrhythmia and occasional extrasystoles do not significantly impair cardiac function, severe tachycardia so shortens the time for diastolic filling of the ventricles and diastolic flow in the coronary arteries that the efficiency of the heart is substantially decreased; this happens in atrial fibrillation and flutter and the paroxysmal tachycardias. The bradycardia seen in complete heart block (about 30 beats a minute) causes a marked fall in cardiac output.

(2) Increased pressure load results from any condition which increases the resistance to expulsion of blood from the ventricles. The commonest causes affecting the left ventricle are systemic hypertension and aortic valve stenosis, while resistance to emptying of the right ventricle is usually due to pulmonary hypertension resulting from various diseases of the lungs and from mitral stenosis.

If the cause is chronic, the affected chamber undergoes hypertrophy, but eventually dilatation and failure develop.

(3) Increased volume load. This arises when a ventricle is required to expel more than the normal volume of blood. It occurs when, owing to incompetence of a heart valve, some of the blood leaks backwards (e.g. through the aortic valve during diastole), and also in conditions in which the general circulation is increased, e.g. anaemia, thyrotoxicosis, and hypoxia resulting from lung disease. Other causes include shunts between the left and right sides of the circulation, and arterio-venous shunts.

(4) Coincidence of multiple factors. Each of the above aetiological groups may independently produce cardiac failure, but various factors often operate simultaneously. For example, a patient with mitral stenosis and myocardial fibrosis due to previous rheumatic fever may have only impaired exercise tolerance (i.e. diminished cardiac reserve) without evidence of cardiac failure at rest; with the onset of atrial fibrillation cardiac failure at rest may develop. An individual with systemic hypertension may develop cardiac failure as a result of occlusion of a minor coronary artery which would go virtually unnoticed but for the hypertension, or cardiac failure may be precipitated by an attack of pneumonia in a person with pulmonary hypertension due to chronic lung disease.

Manifestations of cardiac failure

In mild failure, cardiac function is adequate for the needs of the body at rest, and failure becomes apparent as undue breathlessness on exertion; this is attributable to congestion of the lungs, which stimulates respiratory reflexes. In more severe cardiac failure, cardiac output is inadequate even at rest and structural changes in various organs result. The nature of the changes depends on the duration of the cardiac failure, and on which ventricle predominantly fails.

Acute cardiac failure is due to conditions of sudden onset, e.g. coronary artery occlusion, pulmonary embolism, acute infections, the hypertension of acute glomerulonephritis, the development of an arrhythmia, or rupture of a chamber or valve cusp. When cardiac failure is acute and severe (most often due to myocardial infarction) the acute fall in cardiac output, with consequent diminished tissue perfusion, results in a reaction closely similar to that in hypovolaemic shock, with selective vasoconstriction. The term *cardiogenic shock* (p. 263) is appropriate, but the central venous pressure is raised, and the principles of treatment are quite different: the full picture of shock develops in only a small proportion of cases, but when it does occur the outlook is poor. Conversely, acute heart failure may develop as a result of severe, prolonged hypovolaemic or septic shock (pp. 262–4). In both of these circumstances, if death occurs rapidly the organs behind the failing ventricle show acute venous congestion,

and the distinction between cardiogenic shock and heart failure resulting from shock may depend on the presence or absence of a causal lesion (usually a myocardial infarct) in the heart.

Chronic cardiac failure may follow acute cardiac failure if death or recovery does not occur rapidly. In cardiac disease of gradual onset, e.g. acquired valvular lesions, chronic failure is particularly common. At death the changes of generalised chronic venous congestion (see p. 228) are found, and the heart will show the features of the causative lesion, the dilatation of failure and in some instances compensatory enlargement of the chambers. These are described below.

Left ventricular failure causes 60 per cent of deaths in untreated essential hypertension, and is a common cause of death in myocardial infarction, a condition which mainly affects the left ventricle. It is also often the cause of death in patients with disease of the aortic valve. The clinical and pathological manifestations are predominantly pulmonary, for the left ventricle fails to maintain its output and, in the presence of an efficient right ventricle, blood accumulates in the pulmonary circulation causing congestion in the lungs. Severe breathlessness, cyanosis and pulmonary oedema are the outstanding clinical features. There may be no evidence of right ventricular failure, though some congestion of the neck veins is usual. Paroxysmal nocturnal attacks of dyspnoea are common in systemic hypertensive patients and are due to acute left ventricular failure, perhaps from resorption of peripheral oedema fluid when the patient is recumbent at night. At necropsy there is dilatation of the left ventricle and functional mitral incompetence: acute or chronic venous congestion of the lungs is found with pulmonary oedema and froth in the bronchi.

Right ventricular failure in its most acute form is found in cases of massive pulmonary embolism. The more chronic forms are found in pulmonary arterial hypertension due to lung disease (e.g. fibrosis or chronic bronchitis and emphysema) and to mitral stenosis. In pure right ventricular failure an excess of blood accumulates in the systemic veins, and in addition to the causal lesion there will be dilatation of the right ventricle with functional tricuspid incompetence, distension of the great veins,

e.g. in the root of the neck, and acute or chronic venous congestion of the liver, spleen and kidneys (p. 228 *et seq*). Oedema, ascites, etc., are conspicuous in untreated advanced cases. Right ventricular failure also occurs in mitral stenosis, but there is, in addition, chronic venous congestion of the lungs, with or without pulmonary oedema. The rise in pulmonary venous pressure results in pulmonary arterial hypertension due to increased tone of the pulmonary arterioles.

Total heart failure combines the features of left and right ventricular failure and is found not only in diseases which cause diffuse myocardial damage (e.g. extensive infarcts, myocarditis) but also in states requiring a persistently high cardiac output (thyrotoxicosis, etc.). In addition, when there is left ventricular failure the strain imposed on the right side of the heart by the raised pulmonary pressure sooner or later leads to right ventricular failure.

In some cases of heart failure due to lack of ventricular compliance, there may be little or no ventricular dilatation.

Hypoxic phenomena. When left ventricular output falls suddenly and markedly, the cerebral blood supply is so diminished that the patient loses consciousness. This may be momentary, as in a vaso-vagal attack (p. 260), or it may be rapidly fatal when due to occlusion of a main coronary artery, massive pulmonary embolism or rupture of a cardiac chamber. When the condition is reversible, e.g. during an attack of complete heart block, transient loss of consciousness may occur (*Stokes-Adams attack*); similar attacks occur in patients with incompetence of the aortic valve. In chronic cardiac failure, oxygen deficiency in the tissues is less marked and is seen clinically in the form of cyanosis, sometimes accompanied by mental confusion. Pathologically, the effects of stagnation hypoxia are best seen in tissues such as liver where centrilobular loss of liver cells is usually present in association with venous congestion. A compensatory increase in red cells (erythrocytosis) may occur due to hypoxia.

Cardiac oedema is dealt with on pp. 255–6.

Thrombo-embolic phenomena. Patients with cardiac failure are especially prone to develop deep venous thrombosis in the legs as a result of venous stagnation and muscular inactivity associated with lying in bed: in consequence,

there is a serious risk of pulmonary embolism. Thrombus is also common in the atria and, in some forms of heart failure, in the ventricles: such cardiac thrombi, depending on their site, can give rise to pulmonary or systemic emboli. These complications are described in Chapter 9.

Compensatory enlargement of the heart

When extra work is imposed on the myocardium as a result of a chronic disease (e.g. a valvular lesion or hypertension) **compensatory myocardial hypertrophy** occurs in the walls of the affected chambers and enables them to deal more effectively with the increased work of maintaining the circulation. In conditions of increased volume load, due for example to valvular incompetence, the affected ventricle(s) dilate passively during diastole to contain an increased volume of blood, thus compensating in some degree for the blood which, instead of passing onwards, is regurgitated through the leaking valve. This occurs before the supervention of heart failure, i.e. when ventricular systolic emptying is normal, and has been termed **compensatory dilatation**. Use of the term 'compensatory' for passive over-dilatation is perhaps not well justified, but it does serve to emphasise the difference from the dilatation of ventricular failure, in which emptying is incomplete due to weakness of the myocardial contraction. Both hypertrophy and compensatory dilatation cause cardiac enlargement, which is further increased by the dilatation of cardiac failure.

The limiting factors in cardiac hypertrophy are not understood. A likely one is the difficulty in maintaining adequate myocardial perfusion and respiratory exchange when the diameter of the fibres is increased. Because of the vascular arrangements of the coronary supply, ventricular hypertrophy renders the inner part of the myocardium particularly liable to ischaemia in the presence of coronary artery disease (p. 404).

The assessment of cardiac enlargement. In left ventricular hypertrophy, the weight of the heart is increased above the normal 300–350 g often to over 500 g. Because of its smaller mass, hypertrophy of the right ventricle is only occasionally sufficient to increase markedly the total heart weight. Although hypertrophy of either ventricle is usually obvious at necropsy

from the increased thickness of its wall (Fig. 15.1), the degree of hypertrophy is not readily assessed without taking account of the volume of the chambers, i.e. the degree of dilatation. A more reliable method is to separate and weigh the individual ventricles, making allowance for epicardial fatty tissue and any gross fibrous scars, etc.

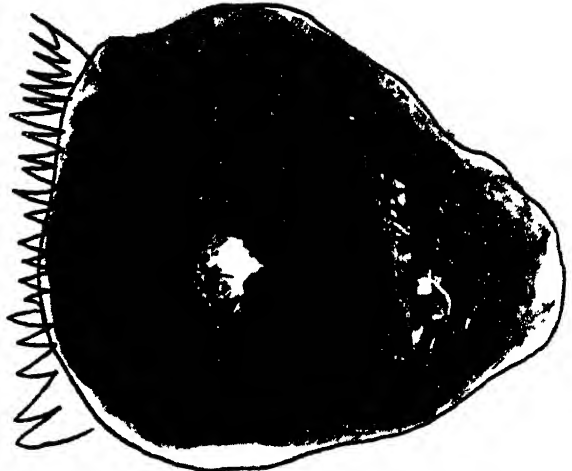


Fig. 15.1 Transverse section through the ventricles of the heart in a case of systemic hypertension, showing obvious hypertrophic thickening of the left ventricle. $\times 0.5$.

Increase in size of the hypertrophied heart is often not readily assessed by clinical or radiological measurements unless there is also dilatation, in which case it is difficult to distinguish, from size alone, between the two processes.

Characteristic electrocardiographic changes accompany ventricular hypertrophy, especially when only one ventricle is involved.

Causes of left ventricular hypertrophy. The common causes of marked left ventricular hypertrophy are (1) systemic hypertension (essential, or secondary to renal disease, coarctation of the aorta and certain endocrine tumours); (2) stenosis of the aortic valve; (3) compensatory dilatation of the left ventricle (in aortic or mitral incompetence); (4) persistently high cardiac output (thyrotoxicosis, anaemia, arterio-venous fistula and Paget's disease of bone). Mild and focal left ventricular hypertrophy is found in the absence of hypertension or valvular lesions in certain patients with healed myocardial infarcts, and is presumably compensatory for the loss of muscle. Various

cardiomyopathies (p. 408) are a less usual cause of hypertrophy.

Causes of right ventricular hypertrophy. Most examples of right ventricular hypertrophy are attributable to pulmonary arterial hypertension. The common causes are (1) chronic lung disease, especially bronchitis and widespread pulmonary fibrosis; (2) stenosis and/or incompetence of the mitral valve; (3) congenital heart disease with large shunts of blood from one side of the heart to the other; (4) stenosis of the pulmonary valve; (5) massive hypertrophy of the left ventricle, which is often accompanied by right ventricular hypertrophy without any other obvious cause. This may be due to distortion of the right ventricular lumen by the hypertrophied ventricular septum. Rarer causes

of right ventricular hypertrophy include multiple small pulmonary emboli and other unusual causes of pulmonary hypertension (pp. 454–7).

Causes of compensatory dilatation. Compensatory dilatation of the left ventricle results from incompetence of the mitral or aortic valve or of both, and right ventricular dilatation from incompetence of the tricuspid and/or pulmonary valves. A lesser degree of compensatory dilatation of both ventricles occurs in patients with persistently high cardiac output, e.g. in thyrotoxicosis or arterio-venous shunts.

Compensatory dilatation is accompanied by hypertrophy of the affected ventricle, the degree of which depends on the increase in its work load.

Ischaemic Heart Disease

Myocardial ischaemia is one of the major causes of disability and death in developed countries, the outstanding cause being atheroma of the coronary arteries, often complicated by occlusive thrombosis. Practically everybody in such communities has some degree of coronary atheroma by the age of 40, and deaths even earlier than this are by no means uncommon. In recent years, however, deaths from ischaemic heart disease have apparently decreased significantly in the U.S.A., perhaps because of the factors mentioned on p. 369. In this country there has been no obvious decrease in the mortality rate, although a plateau may have been reached. Severe atheromatous narrowing, particularly of more than one major coronary artery, can give rise to (1) **angina pectoris**, severe chest pain brought on by factors which increase the work of the heart; (2) **myocardial infarction**, usually precipitated by super-added occlusive thrombosis; (3) **sudden death**; (4) **cardiac failure**; and (5) **cardiac arrhythmias**, due to ischaemic injury of the conducting system.

These last three effects may occur either with or without a history of angina pectoris or myocardial infarction.

Angina pectoris

This consists of attacks of severe, sometimes agonising chest pain of sudden onset, due to

acute ischaemia of a part of the myocardium with an inadequate blood supply. The pain is brought on by factors which increase the work of the heart and is relieved by rest and by vasodilator drugs such as nitroglycerin. Physical exercise, anger, anxiety, a heavy meal and exposure to cold can all induce an attack.

Aetiology and structural changes. The primary factor in nearly all cases is atheromatous narrowing of the coronary arteries. *Coronary atheroma* shows the usual features of atheroma in other arteries (pp. 362–6). There is patchy fibrous thickening of the intima with accumulation of lipid debris (Fig. 15.2). Calcification is usually present, and confluent calcified patches of atheroma may convert lengths of the vessel into a rigid tube which cannot be cut with a knife; it is then very difficult to examine satisfactorily without decalcification. Atheroma affects all the main branches of the coronary arteries but the earliest and most severe lesions often develop in the first 2–3 cm of the left coronary artery or in its anterior descending branch. Even small distal epicardial branches may be narrowed, although branches that have penetrated into the myocardium are usually free of atheroma. At necropsy, occlusion of one or more major coronary vessels by organised thrombus can often be demonstrated, with some restoration of blood supply to the distal part of the occluded vessel by enlarged

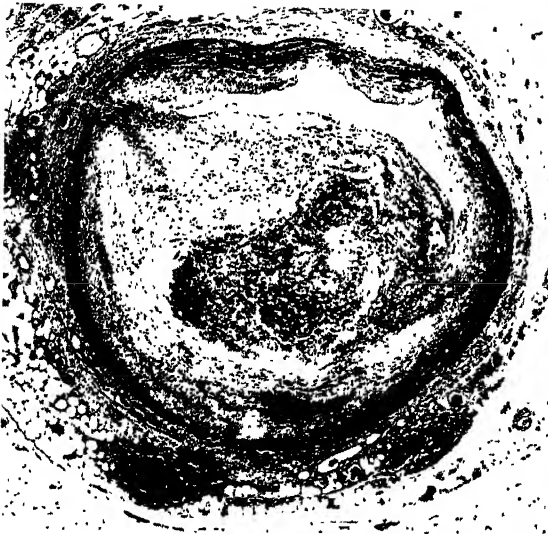


Fig. 15.2 Severe atheromatous narrowing of a coronary artery. In this instance, the lumen has been further reduced by haemorrhage into the soft lipid-rich material, seen as the dark area in the atheromatous patch. $\times 20$.

anastomotic arteries (Fig. 15.3). In most cases of angina pectoris, at the time of death there is myocardial scarring (Fig. 15.4) or recent myocardial infarction, but myocardial lesions are not invariably present.

Severe atheromatous coronary narrowing may account for the finding of areas in which the myocardium is partly replaced by scar tissue, but in which some myocardial fibres persist (Fig. 15.5), particularly around blood vessels. Old, organised coronary occlusions are, however, commonly present in such cases, and it is often not clear whether the scarring has resulted from chronic ischaemia or myocardial infarction. On the basis that the resistance to coronary blood flow is mainly in the arterioles and capillary network, it has been calculated that an atheromatous lesion of a coronary artery must reduce the sectional area by 70 per cent or more to cause a reduction in blood flow. This may be true for resting conditions, but lesser degrees of narrowing, especially if multiple, will inevitably restrict the increased flow which occurs when the arterioles relax, as in physical activity.

In most cases of angina pectoris, there is some degree of left ventricular hypertrophy, usually due to systemic hypertension: by increasing the amount of muscle supplied by the coronary arteries, this aggravates coronary

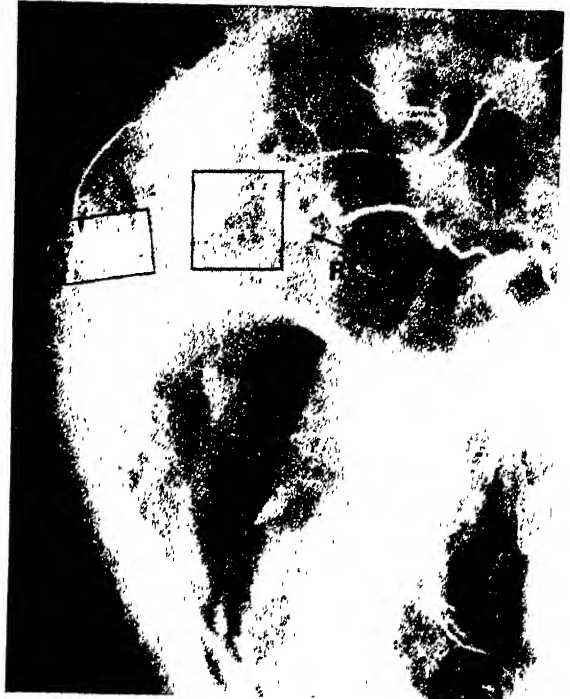


Fig. 15.3 Post-mortem radiograph of part of the right coronary artery in a case of angina. Shortly after its origin (*arrow*), a length of the artery is occluded (*square*) and adjacent vessels have enlarged to provide a collateral route, so that the artery is filled beyond the occlusion. A second occlusion (*oblong*) is present, but is partly obscured by the curving course of the artery. (Professor M. J. Davies.)



Fig. 15.4 Extensive fibrosis of left ventricle with marked thinning at places; secondary to coronary disease. $\times 0.5$.

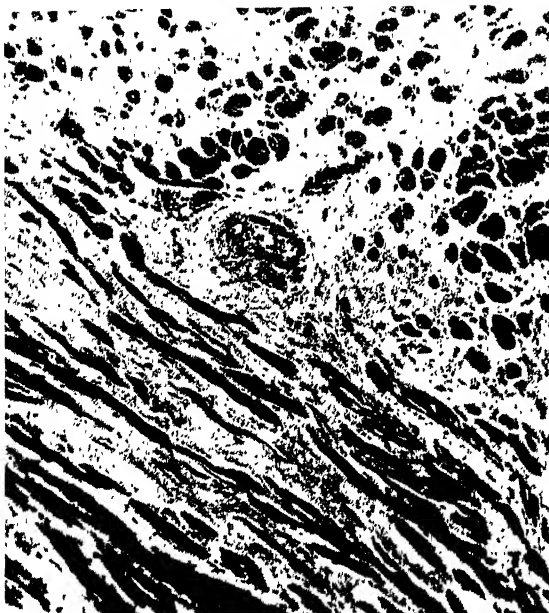


Fig. 15.5 Part of a large zone of myocardial scarring throughout which some myocardial fibres have survived.

insufficiency and acts as a predisposing cause. Ventricular hypertrophy due to lesions of the cardiac valves or to cardiomyopathy (see later) has a similar effect. Narrowing of the ostia of the coronary arteries by syphilitic aortitis is now a rare cause of angina pectoris.

Clinical course. In general, chronic coronary insufficiency can be expected to shorten life, although there are wide variations in the clinical course and some patients with angina survive more than 20 years. Many die suddenly and unexpectedly; in some this is due to sudden occlusion of a coronary artery by thrombus, but in about half the cases no recent occlusion can be demonstrated. Death in these cases is assumed to be due to the sudden onset of ventricular fibrillation, and this has been confirmed in patients being monitored by electrocardiography at the time of death.

Myocardial infarction

Myocardial infarction is the commonest cause of death in many parts of the world today. Approximately 33 per cent of males and 25 per cent of females coming to necropsy in this department have evidence of old or recent myocardial infarction. The mortality rate varies appreciably, even in different parts of the Bri-

tish Isles, and in the Clyde valley the rate is probably the highest in the world.

Like infarction in other tissues, it is due to acute ischaemia, usually caused by occlusive thrombosis of a coronary artery over an atheromatous patch.

Clinical course

The onset of myocardial infarction is signalled by the appearance, often abruptly, of severe persistent chest pain. Many patients already suffer from angina pectoris and they often recognise that the pain is different, being continuous for some hours and failing to respond to rest and nitroglycerin. With the onset the patient may experience profound weakness and breathlessness, and there is usually evidence of peripheral circulatory failure with hypotension, cyanosis and a cold clammy skin. Fever, leucocytosis and a raised erythrocyte sedimentation rate are commonly observed within the first 24 hours. Certain tissue enzymes are released by the dead heart muscle, with a consequent rise in blood levels, which is of diagnostic value. The serum concentration of glutamic oxaloacetic aminotransferase, for example, rises within 6 to 12 hours, and reaches a peak within 2 or 3 days. Characteristic abnormalities of the electrocardiogram are often demonstrable, but by no means always. The prognosis of myocardial infarction depends on a number of factors including the size of the infarct, its site (e.g. whether it involves the conducting system), the patient's age, and whether or not he (or she) has hypertension. Approximately 25 per cent of the patients admitted to hospital die, mostly within the first week, but this excludes a large number who die before they can reach hospital. Death is most commonly caused by ventricular fibrillation, cardiac failure or secondary embolic disease (see below).

Structural changes

The changes found in the heart at necropsy depend on how long the patient has lived since the onset of infarction.

The **coronary arteries** usually show extensive atheroma, but in some cases there will be only one or two patches causing severe narrowing. The occluding thrombus typically overlies an atheromatous patch, but may extend along a

considerable length of the vessel. At first, it is usually composed mainly of red thrombus (Fig. 15.6); subsequently it becomes pale and is eventually organised, often with some recanalisation (Fig. 15.7).

Thrombosis of the anterior descending branch of the left coronary artery is particularly common and usually causes an infarct involving the anterior wall of the left ventricle together with the apex and sometimes extending to the anterior part of the interventricular septum and the adjacent part of the anterior



Fig. 15.6 Recent thrombotic occlusion of the right main coronary artery: death resulted 12 hours after the onset of symptoms. The artery has been opened longitudinally to reveal the thrombus.



Fig. 15.7 Coronary artery recanalised after occlusion by thrombus. There was extensive healed myocardial infarction. $\times 25$.



Fig. 15.8 Post-mortem angiogram in a case of recent myocardial infarction. The anterior descending artery (LAD) is occluded proximally (arrows), and although there was sufficient collateral supply to fill the artery distally, extensive infarction had occurred over the anterior wall and apical region. Note also the atheromatous narrowings of the right coronary artery. (Professor M. J. Davies.)

wall of the right ventricle (Fig. 15.8). Occlusion of the right main coronary artery is almost as frequent and is associated with infarction of the posterior wall of the *left* ventricle, sometimes involving also the posterior part of the septum and posterior wall of the right ventricle. Occlusion of the circumflex branch of the left coronary artery can cause infarction of the lateral wall of the left ventricle, but occlusion of the descending and circumflex branches of the right coronary artery are relatively rare causes of infarction. In some instances, recent thrombus is found in two major vessels, and infarction is accordingly extensive.

The extent of infarction varies considerably, depending on the severity of atheromatous narrowing and the presence or absence of previous thrombosis in other coronary arteries. There is considerable normal variation in the relative sizes of the left and right arteries, and atheromatous narrowing of a branch may result in enlargement of collateral vessels (Fig.

15.3) so that its final occlusion causes a lesser infarct than usual, or even none at all.

The detection of thrombus in coronary arteries narrowed by atheroma is not always easy. If the arteries are opened lengthwise with scissors, a small thrombus may be displaced and disrupted by the point of the blade without being noticed. A more satisfactory method of examining the vessels is by transverse section at close intervals. The presence of extensive and grossly calcified atheromatous lesions prevents satisfactory naked-eye examination of the affected parts of the vessels at necropsy, and decalcification is first necessary.

The **infarct** may involve the whole thickness of the myocardium, or it may be confined to the inner part of the wall. Occasionally infarction involves the deeper, subendocardial myocardium throughout most of the circumference of the left ventricle: there is usually severe atheromatous narrowing of the left and right main coronary arteries and the anterior descending branch, but without recent thrombosis. It appears that such extensive arterial narrowing renders especially precarious the blood supply to this part of the myocardium and that subendocardial ischaemic necrosis can occur without superadded occlusive arterial thrombosis (Davies, Woolf and Robertson, 1976).

Although ischaemic death of myocardium occurs within a few minutes of loss of blood supply, the visible changes of infarction in the dead muscle do not appear for 8 hours or so. Accordingly, in patients who die suddenly at the time of coronary occlusion, or within the next few hours, acute changes in the myocardium are not observed by naked-eye or ordinary light microscopy. The first changes noticed by naked eye are blotchy congestion and pallor. Although the muscle undergoes coagulative necrosis, the infarct can usually be felt after a day or so as a patch of softening, and during the next few days it becomes pale (Fig. 15.9) and its colour changes from brown to yellowish-grey. After a week or so it becomes more sharply defined by the development of a zone of vascular granulation tissue along the margin, and removal of the dead tissue by organisation gradually proceeds. When the infarct extends to the outer surface of the myocardium, the pericardial surface is often covered by a layer of fibrinous exudate with marginal haemorrhages. On the inner

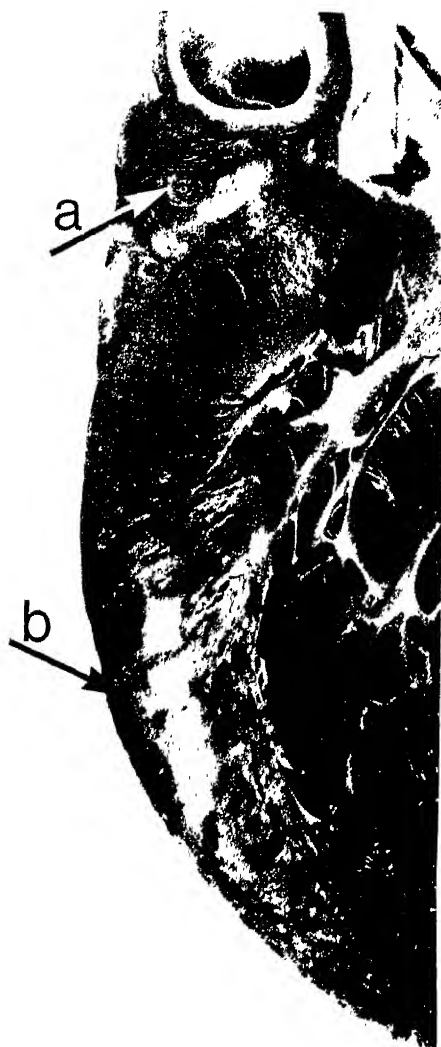


Fig. 15.9 Myocardial infarction of 11 days' duration. The anterior descending branch of the left coronary artery (**a**) is occluded by thrombus and there is extensive infarction of the wall of the left ventricle, seen as areas of pallor and surrounding congestion (**b**). Note also mural thrombus at the apex.

aspect, the endocardium and a thin layer of adjacent myocardium remain alive, nourished by blood from the lumen, but in patients surviving for several days this does not prevent thrombosis on the endocardial surface (see below).

Under the microscope the infarcted muscle shows the usual features of necrosis (Fig. 2.5, p. 11). The necrotic muscle is invaded by polymorphonuclear leukocytes and, after a few days, digestion by macrophages and organisation can be seen at the margins (Fig. 15.10).

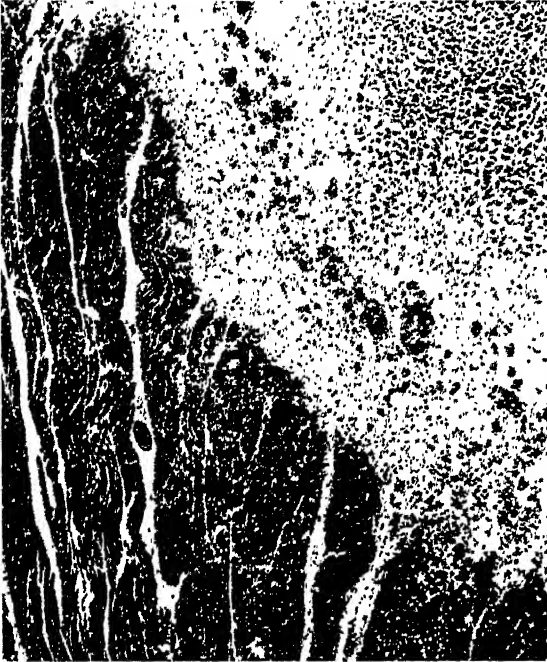


Fig. 15.10 Infarct of myocardium of 12 days' duration. The necrotic heart muscle (*upper right*) is separated from the surviving muscle by a zone of cellular and vascular granulation tissue. $\times 40$.

Gradually, the dead muscle is replaced by fibrous tissue, the process taking some weeks or even months, depending on the size of the infarct, and eventually a pale fibrous scar remains (Fig. 15.11).

As already stated, it is sometimes not possible to determine whether myocardial fibrosis has resulted from infarction or chronic ischaemia.



Fig. 15.11 Old infarct represented by replacement of the lateral and posterior wall of the left ventricle by a relatively thin fibrous scar.

Aetiology

The important cause of myocardial infarction is atheroma of the coronary arteries with super-added occlusive thrombosis over an atheromatous patch. The aetiology of atheroma, and the factors associated with an increased risk of myocardial infarction, have been discussed on p. 366. These include raised levels of various plasma lipids, obesity, hypertension, cigarette smoking and a sedentary occupation. These factors are all of long duration, and are likely to promote the gradual development and extension of atheroma, but they may also predispose to superadded thrombosis. Enhanced coagulability of the blood has been shown to be common in patients with ischaemic heart disease, and correlates partially with the levels of plasma lipids. It has also been shown that coronary blood flow is affected by cigarette smoking, which may also predispose to thrombosis by the mechanisms noted on p. 369.

Patients with severe coronary atheroma are particularly liable to myocardial infarction following a severe injury or major surgical operation. This may occur during a period of shock, but the risk is high for some weeks following the injury, etc. The circulatory disturbances of shock, the effect of anaesthetic agents on the heart, and the increased coagulability of the blood following injury are all probably of importance.

Relationship of coronary thrombosis to myocardial infarction. The reported incidence of recent coronary thrombosis in patients dying of acute myocardial infarction varies greatly. In some series, the incidence has been only about 50 per cent, and considerably lower in those patients dying within a few hours of the onset of symptoms. Accordingly, it has been suggested that coronary thrombosis may not be the cause of infarction, and that it occurs *after* the muscle has died. The evidence for this view depends upon the care with which the coronary arteries are examined (see above). By careful examination of the coronary arteries by close transverse section, preceded when necessary by decalcification, workers at St. Georges' Hospital, London, have found an occlusive thrombus in the expected artery in virtually all cases of myocardial infarction in patients dying 12 hours or more after the onset of clinical symptoms, and Harland's experience in this depart-

ment has been similar. In patients dying within a few hours of onset, infarction cannot be identified morphologically, and the diagnosis is insecure. These findings relate to regional (transmural) infarcts occurring mainly within the territory supplied by a major coronary artery. By contrast, circumferential sub-endocardial necrosis of the myocardium (see above) is not usually associated with recent coronary thrombosis, but this lesion accounts for only about 6 per cent of all infarcts (Davies, Woolf and Robertson, 1976).

Microscopic examination of recent coronary occlusions shows that the thrombus very often forms over a *ruptured* atheromatous plaque (Fig. 15.12), and this suggests that the rupture (and not necrosis of muscle) has promoted thrombosis. It follows that, in such cases, infarction is the *result* of the occlusion. Admittedly, old organised occlusive coronary arterial

thrombosis is sometimes found without evidence of previous infarction (i.e. a myocardial scar), but this is probably attributable to compensatory enlargement of anastomotic arteries due to the presence of atheroma before the thrombosis occurred.

Experimental animal studies also support the causal role of thrombosis, for occlusion of a major coronary artery results in infarction, whereas myocardial necrosis induced by other means is not followed by coronary thrombosis.

In his extensive studies on myocardial infarction, Fulton has found no evidence against the orthodox view that coronary artery thrombosis precedes and causes transmural infarction. In particular, the evidence from examination of coronary thrombi at necropsy following intravenous administration of radio-labelled fibrinogen was fully consistent with this view (see Davies, Fulton and Robertson, 1979).

Infrequent causes of myocardial infarction. Occlusion of the ostia of the coronary arteries by syphilitic aortitis is nowadays a rare cause of infarction, and so are coronary lesions due to polyarteritis nodosa or Buerger's disease. Occlusion of a coronary artery by embolus is also much less common than thrombosis, but sometimes results from bacterial endocarditis, and also from detachment of thrombus forming in relation to valve prostheses.

Effects and complications

(a) **Arrhythmias.** Death from myocardial infarction most frequently results from arrhythmias, particularly ventricular fibrillation. Healed infarcts may also be associated with cardiac arrhythmias and are the chief cause of heart block.

(b) **Cardiac failure.** Extensive infarction of left ventricular muscle can cause acute heart failure, and if severe this results in **cardiogenic shock**, the prognosis of which is poor. Loss of the infarcted muscle also predisposes to chronic heart failure, which may develop at any time after infarction.

(c) **Mural thrombosis.** Following acute myocardial infarction, release of tissue thromboplastin from the damaged muscle and localised eddying of blood may lead to thrombosis in the ventricular chambers (Fig. 9.16, p. 238). This is seen at necropsy in about 30 per cent of cases: in patients who survive, the thrombus is eventually



Fig. 15.12 Coronary artery in longitudinal section, showing an ulcerated atheromatous plaque (left, lower) with occlusion of lumen by thrombus. $\times 15$.

organised. Systemic emboli can result from mural thrombosis, but are not very common.

(d) Venous thrombosis. Presumably because of reduced blood flow, systemic venous thrombosis is an important complication of myocardial infarction and tends to occur especially in the veins of the legs (p. 241). Detachment of such thrombus is common and consequently pulmonary embolism is not infrequently a cause of death in myocardial infarction.

(e) Rupture of the left ventricle due to myocardial softening (*myomalacia cordis*) causes 10 per cent of deaths occurring soon after myocardial infarction. It generally occurs within 14 days after infarction, when autolysis of the infarct is active, and especially when leukocytic invasion is marked and repair has not started. Rupture leads to sudden death from massive haemopericardium. When the interventricular septum is involved in infarction it may rupture, leading to sudden onset of severe cardiac failure and a loud heart murmur. Rupture of an infarcted papillary muscle in the left ventricle may also occur, leading to incompetence of the mitral valve, with a loud murmur and intractable pulmonary oedema. Persistence of hypertension and of physical activity after infarction predispose to cardiac rupture.

(f) Cardiac aneurysm. Occasionally the fibrosed wall of a healed infarct of the left ventricle may stretch to form a cardiac aneurysm. As with other aneurysms, laminated thrombus tends to form in the cavity (Fig. 15.13).

(g) Angina pectoris. Whenever myocardial infarction has occurred, the adjacent myocardium, although not infarcted, is likely to be ischaemic (and is presumably the source of the prolonged pain associated with infarction). As anastomotic channels dilate and enlarge, the blood supply to such areas of partial ischaemia will improve. However, in some patients angina pectoris dates from a myocardial infarction, and it is apparent that thrombosis of a major coronary vessel may render areas of myocardium chronically ischaemic. In some instances, angina is cured by myocardial infarction, presumably because an area of myocardium which was previously chronically ischaemic has been included in the infarct and destroyed.

(h) Recurrence of infarction. Because atheroma is generally extensive, individuals who have had a myocardial infarct are prone to recurrence.



Fig. 15.13 Aneurysm of the left ventricle following ischaemic myocardial fibrosis. Much of the anterior wall (*left*) of the ventricle has been replaced by fibrous tissue, which has stretched to form an aneurysm, now largely filled with laminated thrombus. $\times 0.7$.

Other effects of ischaemic heart disease

Ischaemic heart disease is the usual cause of **sudden death**. There may be a history of angina, previous infarction, evidence of chronic heart failure, or chest pain immediately before death, but sometimes there have been no warning symptoms. At necropsy, there is usually severe coronary atheroma, with or without old organised thrombotic occlusions. In approximately half of these cases, there is a recent occlusive thrombus or a ruptured atheromatous plaque which has apparently caused death before there has been time for the morphological changes of infarction to develop. The occasional unexpected deaths which have occurred during ECG recording suggest that ventricular fibrillation is the usual cause.

Ischaemic heart disease is also the most important cause of **chronic heart failure** and of **cardiac arrhythmias**, whether or not there has been previous myocardial infarction.

Non-inflammatory, Non-ischaemic Disorders of the Myocardium

Fatty change of the myocardium (p. 24) is an expression of impaired metabolism of the heart muscle most frequently seen in severe prolonged anaemia. *Adiposity of the heart* is a deposition of fat in the epicardium which extends between cardiac muscle fibres (Fig. 2.19, p. 28). It affects the right ventricle especially and is usually, but not always, associated with general obesity. Very rarely cardiac function is impaired by this deposition of adipose tissue. *Brown atrophy* (p. 39) is not related to the occurrence of cardiac failure. *Primary amyloid disease* (p. 272) commonly affects the heart and is a rare cause of cardiac failure. The clinical picture may be similar to that of restrictive cardiomyopathy, and diagnosis may be difficult both clinically and at necropsy unless the possibility of amyloidosis is kept in mind. *Segmentation and fragmentation of the myocardium* indicate conditions in which the heart muscle cells either separate from one another at the intercalated discs or show irregular tearing. These changes have been observed in cases of violent death. It is supposed that they result from irregular contraction at the time of death.

Cardiomyopathy

Cardiomyopathy is a convenient term for the classification of a heterogeneous group of chronic myocardial disorders which are not due to ischaemia, hypertension, valve disease or shunts, and appear non-inflammatory. Inclusion of ischaemic myocardial disease and various forms of myocarditis has also been proposed, but the term would then apply to nearly all myocardial lesions and would be of little value.

The four generally accepted types of cardiomyopathy are as follows.

1. Hypertrophic cardiomyopathy with or without obstruction. In this condition there is asymmetrical hypertrophy of the left ventricle, and especially of the septum. Function is affected by (a) undue rigidity of the left ventricle, which interferes with diastolic filling, and (b) in many, but not all cases, the bulging, hypertrophied septum obstructs the outflow

of the left, and less commonly the right, ventricle.

Clinically, the condition presents from early childhood to old age, usually as heart failure or atypical angina: it also causes sudden death. It is often familial and has been detected (by echo cardiography) in some apparently healthy relatives, inheritance apparently being determined by a Mendelian dominant factor. Microscopy shows interstitial fibrosis, areas of disordered, whorled arrangement of muscle fibres and very marked thickening of the individual fibres, with enlarged, pleomorphic nuclei and increased glycogen content. These features are useful in diagnostic biopsy. The nature of the basic abnormality is unknown and current views are largely speculative.

2. Congestive cardiomyopathy. This consists of congestive heart failure which is not due to any of the known causes, and is thus diagnosed by exclusion. At necropsy all the chambers are dilated, the myocardium is pale and unduly flabby, and there is often ventricular mural thrombus and endocardial thickening. Microscopy may show gross hypertrophy of some fibres, with conspicuous nuclear enlargement, and atrophy of others. There may also be interstitial fibrosis. These changes are not diagnostic.

Congestive cardiomyopathy sometimes shows a familial tendency, and may be associated with alcoholism or follow childbirth, but the causation is entirely unknown and it occurs from childhood to old age. Congestive failure is a complication of various other conditions, e.g. dystrophy of the skeletal muscles, Friedreich's ataxia, glycogen disease of the Pompe type, beri-beri and acromegaly. Such cases are sometimes included as congestive cardiomyopathy but the cause, if known, should be specified.

3. Restrictive cardiomyopathy is also known as **endomyocardial fibro-elastosis**. It is a rare disorder, mainly of infants, characterised by a thick smooth layer of collagenous and elastic tissue between the endocardium and the myocardium. Involvement of the left ventricle, which may be hypertrophied, is most common. Cases have been attributed to fetal endocarditis, to anoxia due to origin of the left coronary artery

from the pulmonary trunk, and to hypoplasia of the left ventricle associated with premature closure of the foramen ovale. Although not really a disease of the myocardium, it is a cause of heart failure of cryptic origin, and so is grouped with the cardiomyopathies.

Amyloid disease of the heart can cause a restrictive cardiomyopathy in adults.

4. Obliterative cardiomyopathy (endomyocardial fibrosis) is of unknown cause and found mainly in tropical Africa where it is one of the common forms of left or total heart failure. It occurs in many other parts of the world. In the absence of significantly narrowed coronary arteries there is dense fibrosis of the endocardium affecting usually the apex and

posterior walls of one or both ventricles. Sometimes the fibrosis partially obliterates the ventricular cavity and, by enveloping the posterior papillary muscles and chordae tendineae, distorts the posterior cusps of the mitral and tricuspid valves with resulting incompetence. Mural thrombus overlying the fibrotic endocardium is common though embolism seldom occurs. The fibrosis is dense and acellular on the surface but more loose with some inflammatory reaction in the deeper layers. Fibrous tissue bands may spread through the inner third of the myocardial wall and there may be atrophy of myocardial fibres with loss of sarcoplasm. Bacterial endocarditis is recorded in about 10 per cent of fatal cases.

Inflammatory Lesions of the Heart

Myocarditis

The term myocarditis is used loosely to cover various lesions of diverse nature, some of which are due to bacterial and viral infections, others to the effects of bacterial toxins. Rheumatic myocarditis appears to be a hypersensitivity reaction to streptococcal antigens. The subject is difficult because of the impossibility of confirming myocarditis in suspected cases which recover, and the poor correlation between pathological lesions in the myocardium and clinical evidence of heart disease. The possibility of serious metabolic disorder with no histological change in the myocardium, and the difficulty of interpreting the results of bacteriological cultures of necropsy material, further complicate the subject.

In fatal cases of acute myocarditis, the heart is flabby, usually pale, the ventricles are dilated and there may be mural thrombosis.

Toxic myocarditis is a major feature of diphtheria. Similar appearances, presumed to be toxic in origin, may be seen in pneumococcal pneumonia, typhoid fever, septicaemia and other extensive septic conditions.

Morphological changes. The gross changes, mentioned above, are not diagnostic. Microscopically there is a parenchymatous lesion with numerous small foci of coagulative necrosis in the muscle. The affected fibres appear

swollen and glassy, with loss of striations and nuclei, and around them there is infiltration, mostly of macrophages and lymphocytes, but polymorphs also may be present. The necrotic fibres afterwards undergo absorption (Fig. 7.1, p. 178), while the supporting cells in the areas of infiltration proliferate, and small fibrous patches ultimately result. The nature of the infection cannot be deduced from the appearances of the cardiac lesions. In some cases of toxic myocarditis due to diphtheria, the conducting system is severely affected (Fig. 15.14), with resultant heart block.

Clinically toxic myocarditis is recognised by the onset of cardiac arrhythmia or acute cardiac failure in a patient with diphtheria, pneumonia or other toxic infection. It may cause sudden death. Peripheral circulatory failure may also be present in severe cases.

Suppurative myocarditis is due to direct extension of pyogenic organisms from an adjacent valve in bacterial endocarditis, or to infection by the bloodstream. The myocardium is a common site of multiple abscess formation in septicaemia or pyaemia, particularly when due to *Staph. aureus*. In pyaemia, the abscesses have the usual haemorrhagic margin, and their presence is suggested at necropsy by small epicardial haemorrhages, incision of which may reveal underlying abscesses. Septic emboli may be found in coronary branches (Fig. 15.15).



Fig. 15.14 Necrosis of the fibres of the left bundle branch, with an inflammatory reaction, in a fatal case of diphtheria. $\times 225$. (Professor A. C. Lendrum.)



Fig. 15.15 Septic embolus in coronary artery in acute infective endocarditis. $\times 36$.

Involvement of the myocardium in acute infective endocarditis is described on p. 423.

Virus myocarditis. Coxsackie viruses of Groups A and B cause myocarditis and pericarditis: men are most often affected but outbreaks in infant nurseries have been described. The myocardium shows widespread damage of the muscle fibres with abundant macrophages, lymphocytes, plasma cells and eosinophils in the interstitial tissue (Fig. 15.16).

Isolated myocarditis. Various forms of sub-acute and chronic myocarditis without con-



Fig. 15.16 Myocarditis due to Coxsackie B virus from a child of 11 months. The field illustrates the extensive focal infiltration with macrophages, lymphocytes, etc. $\times 320$. (Dr. J. F. Boyd.)

comitant endocarditis or pericarditis have been described, and although the appearances resemble those in Coxsackie infection in the newborn, nothing is known of their causes or relationships. In one variety, known as *interstitial* or *Fiedler's myocarditis*, the heart is dilated and hypertrophied and mural thrombi are common. Yellowish-white foci of necrosis may be visible and microscopically these show conspicuous infiltration of the interstitial tissue around the necrotic muscle fibres with macrophages, lymphocytes, plasma cells, eosinophils and also multinucleated giant cells apparently derived from the damaged muscle fibres. Figure 15.17 is from a boy aged 14 who characteristically died suddenly after a brief illness.

Clinically the disease presents with cardiac arrhythmias, chest pain and embolic phenomena, progressing in a few weeks or months to cardiac failure without obvious cause.

Sarcoidosis may involve the heart, granulomas developing in the myocardium. The endocardium and pericardium are not usually involved.

Syphilis was formerly a frequent cause of

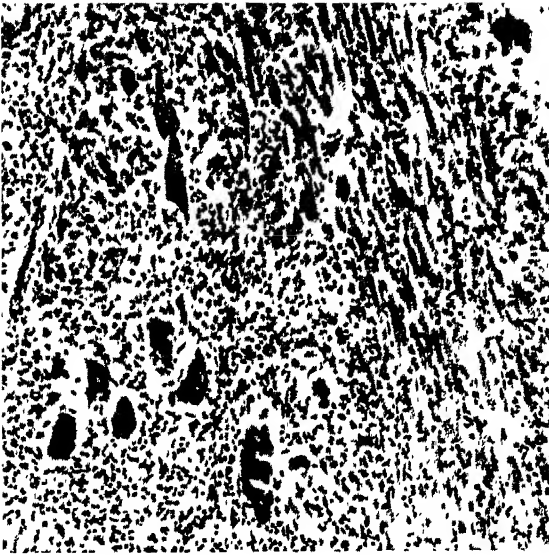


Fig. 15.17 Acute interstitial myocarditis. The heart muscle fibres are replaced by a granulomatous reaction, in which there are many giant cells derived from the muscle fibres. $\times 85$. (Professor Sir Tom Symington.)

serious heart disease. It may affect the heart in four ways. *Incompetence of the aortic valve* is the commonest form and is due to stretching of the valve ring as a result of its involvement in syphilitic mesaortitis. Because they do not meet, the cusps lose their mutual support during diastole and so become elongated and tend to sag (Fig. 15.18). The effects on the chambers of the heart are similar to those resulting from any other variety of aortic incompetence (see p. 419).

Myocardial ischaemia may follow narrowing of the orifices of the coronary arteries by intimal fibrous plaques formed over areas of mesaortitis or by cicatricial contraction of healing lesions. Angina pectoris, myocardial infarction and sudden death are, however, very rarely due to syphilis nowadays.

Gumma of the heart is very rare. It may involve the conducting system and cause heart-block. In *congenital syphilis*, there may be miliary gummas (Fig. 15.19) or interstitial fibrosis of the myocardium.



Fig. 15.18 Syphilitic aortitis, affecting the origin of the aorta. Stretching of the aortic ring has resulted in sagging and thickening of the cusps: dilatation and hypertrophy of the left ventricle are secondary to aortic incompetence. $\times 0.5$.

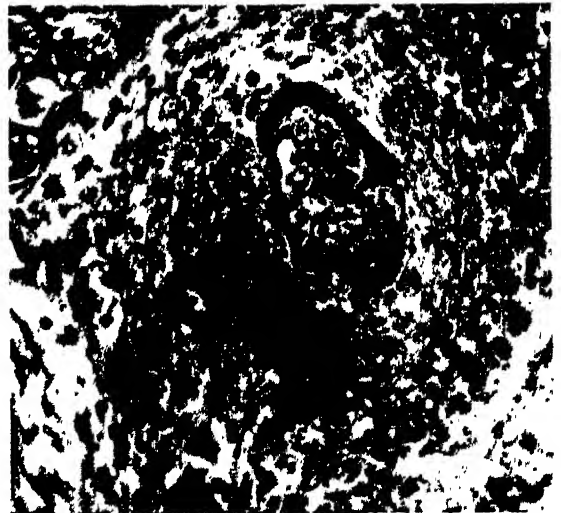


Fig. 15.19 A miliary gumma in the myocardium of a child with congenital syphilis. There is a patch of necrosis adjacent to a small thrombosed artery, and the usual surrounding infiltrate of lymphocytes, plasma cells and macrophages. $\times 300$. (The late Professor J. W. S. Blacklock.)

Rheumatic heart disease

Acute rheumatic fever

This is an acute febrile illness in which lesions occur in the heart, the joints and the subcutaneous tissue. It follows an attack of streptococcal pharyngitis and occurs mainly in children and young adults. Its incidence has fallen greatly in developed countries as a result of improved living conditions and use of antibiotics, but it is common in some tropical countries, including parts of Africa and India.

The function of the myocardium is seriously affected in the acute illness, but recovery of the myocardium is usually complete and any residual effects are due to damage to the heart valves, which may become permanently deformed with consequent increase in the workload of the heart and eventually chronic heart failure. Rheumatic fever shows a marked tendency to recur with subsequent attacks of pharyngitis, and increasingly severe valvular disease may result.

Changes in the heart

Most patients recover from acute rheumatic fever but occasionally it causes death from acute myocardial failure. This is sometimes precipitated by the additional load of acute pericarditis on the weakened myocardium, particularly if there is much pericardial effusion.

The **pericarditis** is exudative, with effusion of fluid and deposition of fibrin on both visceral and parietal layers: the surfaces have a roughened appearance (Fig. 3.21, p. 66) which has been likened to that produced by pulling apart two pieces of buttered bread. Unless very scanty, the fibrin is subsequently removed by organisation, which results in fibrous thickening of the pericardium and often adhesion between the two layers with partial or complete obliteration of the sac.

In fatal acute cases, the **myocardium** is flabby and the ventricles, particularly the left, are dilated. Apart from this, there are no obvious naked-eye myocardial changes, but sometimes tiny pale foci may be just visible: these are the **Aschoff bodies** which are pathognomonic of rheumatic carditis. They are scattered throughout the myocardium (Fig. 15.20), being par-

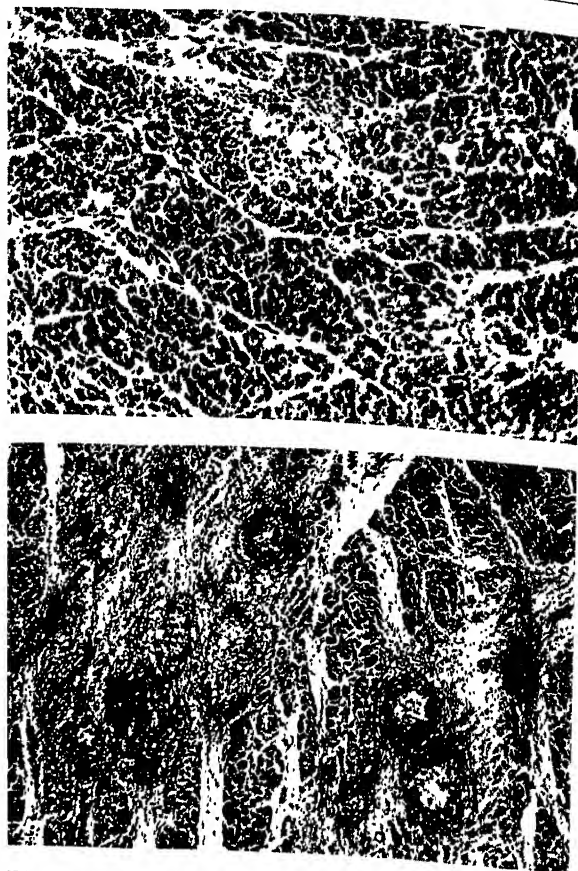


Fig. 15.20 Rheumatic myocarditis. *Above*, the acute stage, showing several Aschoff bodies. $\times 80$. *Below*, later stage, showing fibrosis around the Aschoff bodies (Masson's trichrome stain: collagen appears dark). $\times 48$.

ticularly numerous in the left atrium and ventricle. The Aschoff body appears to be centred on the strands of connective tissue which run through the myocardium, and microscopy reveals a focus of eosinophilic hyaline change in collagen fibres, surrounded by an aggregate of lymphocytes, macrophages, occasional polymorphs, and larger cells with two or three nuclei or a single convoluted nucleus (Fig. 15.21). In time, the Aschoff bodies subside and healing occurs with fibrosis, leaving minute focal scarring in the connective tissue of the myocardium.

Although the function of the myocardium is seriously impaired in rheumatic fever, diffuse myocardial changes in fatal cases are neither characteristic nor impressive. There is some

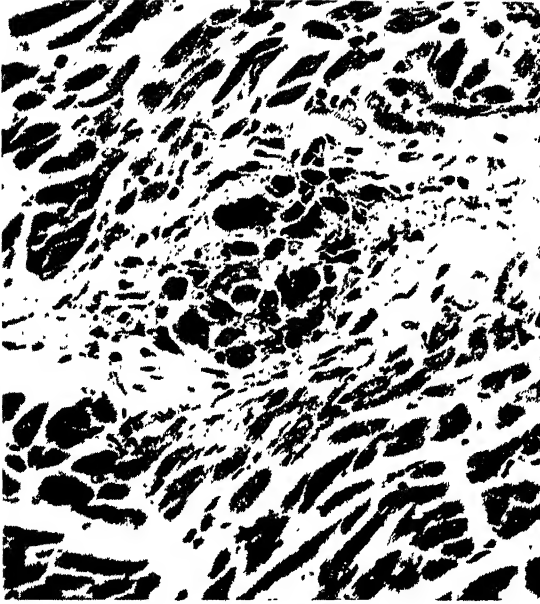


Fig. 15.21 Aschoff body in the myocardium of a child who died of heart failure during an attack of acute rheumatic fever. Central hyaline material is surrounded by macrophages, some with large or multiple nuclei, and by lymphocytes, etc. $\times 220$.

inflammatory oedema and a slight scattering of lymphocytes and occasional polymorphs. The functional disturbance appears to be due to an immunological reaction (see below).

The **endocardium** (like the myocardium) shows diffuse inflammatory oedema and light cellular infiltration. Aschoff bodies also de-

velop in the endocardium and are particularly numerous in the posterior wall of the left atrium just above the insertion of the posterior mitral cusp (McCallum's area); their healing may result in thickening and irregularity of the endocardium in this area. In the acute fever, however, the most prominent endocardial lesion is seen on the heart valves and consists of small **thrombotic vegetations**, forming an interrupted or continuous line of fine grey-pink nodular deposits on the surface of the valve cusps. The vegetations consist at first mainly of platelets, on which fibrin is later deposited (Fig. 9.14, p. 237). They form on that part of each valve cusp which comes into contact with the opposing cusp when the valves close, and are thus seen near the free margins of the cusps, on the atrial surface of the mitral (Fig. 15.22) and the ventricular surface of the aortic valve. Vegetations develop most commonly on both these valves, but quite often on the mitral valve alone and less commonly on the aortic valve alone. The tricuspid also is occasionally affected, but very rarely the pulmonary valve.

Microscopically, the vegetations appear eosinophilic and refractile (Fig. 15.23): the underlying tissue of the cusp shows inflammatory oedema, superficial ulceration, infiltration with lymphocytes, plasma cells, macrophages and polymorphs and proliferation of fibroblasts. In their inflamed oedematous condition, the impact of closure and



Fig. 15.22 Mitral valve in acute rheumatic endocarditis, showing the small vegetations which form along the line of apposition of the cusps. $\times 2$.

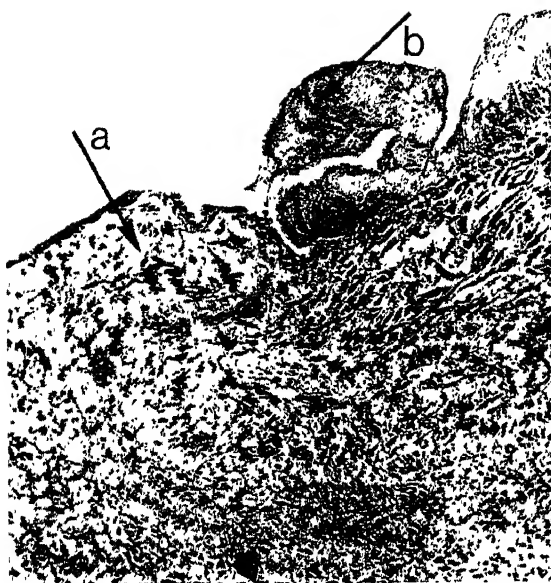


Fig. 15.23 Section of the mitral valve in acute rheumatic endocarditis. The cusp shows inflammatory oedema (a) and cellular infiltration, and appears to be vascularised. Where the cusps meet, the oedematous tissue has ulcerated and platelets have been deposited on the ulcerated surface to form the early vegetation (b). Subsequently, fibrin also is deposited (see also Fig. 9.14, p. 237). $\times 80$. (Professor A. C. Lendrum.)

mutual pressure during the closed state appears to injure the cusps sufficiently to result in loss of endothelium and damage to the underlying connective tissue: in consequence, platelets and fibrin are deposited on the injured surface. The vegetations do not seriously impair valvular function at this stage, although they tend to glue together the margins of adjacent cusps close to the commissures. They are firmly adherent to the cusps and do not produce emboli. Vegetations may form also on the chordae tendineae, particularly of the mitral valve, and in McCallum's area of the left atrium. The tissue at the base of the mitral and aortic valve cusps is often infiltrated heavily with lymphocytes, macrophages, etc., but without the focal arrangement of the Aschoff body.

These changes are followed by vascularisation of the (normally almost avascular) cusps of the affected valves. Blood vessels around the base extend up into the cusps towards the free margin, and the vegetations are removed by the process of organisation, resulting in fibrous thickening and adherence of the cusps near the free margin. Vascularisation is also accom-

panied by more general fibrosis, so that the cusps are irregularly thickened and distorted. Fibrosis occurs also in the papillary muscles and chordae: the latter become thicker and shorter and sometimes bound together by fibrous tissue resulting from organisation of vegetations on them. With recurrent attacks, the acute changes are superimposed on the damaged tissues and distortion of the valves, described more fully on p. 417, becomes progressively more severe.

Micro-organisms are not detected in the heart in rheumatic fever.

Changes in other tissues

In patients dying of acute rheumatic fever, the lungs are congested, heavy, and feel firm and rubbery. Microscopy shows acute congestion, accumulation of oedema fluid containing some macrophages and desquamated pneumocytes, and lining of the alveolar ducts by a dense layer of fibrin ('hyaline-membrane disease', Fig. 15.24). These changes were formerly regarded

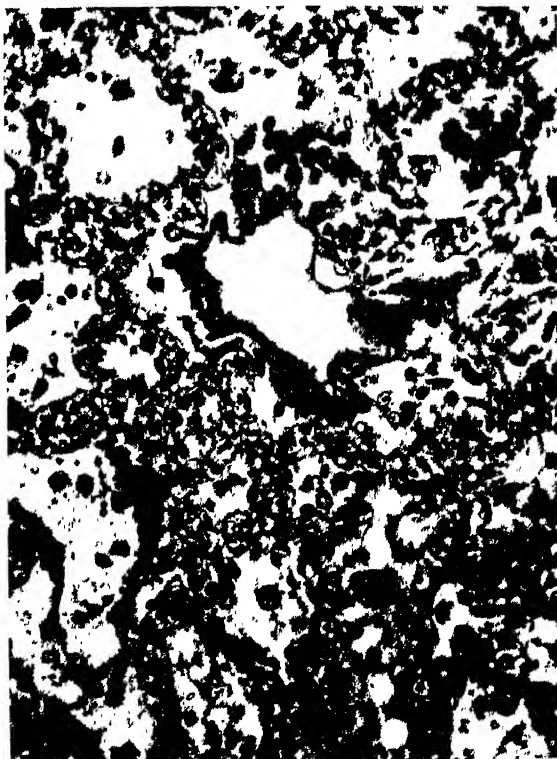


Fig. 15.24 The lung in a case of cardiac failure due to acute rheumatic myocarditis. The lung shows hyaline membranes lining alveolar ducts and mononuclear cells in the alveolar walls and lying free in the alveoli. $\times 220$.

as a specific rheumatic pneumonia, but they appear to result from fairly acute left ventricular failure and are sometimes seen when this occurs from causes other than rheumatic fever. The **joints** show mild inflammatory changes in the synovium with cellular infiltrates resembling Aschoff bodies but more diffuse; the tendons and their sheaths may show similar changes.

In some cases, **subcutaneous nodules** develop over bony prominences of the arms and legs, the commonest sites being the extensor surface of the elbow and overlying the ulna. The nodules are usually between 1 and 2 cm in diameter, painless, and consist of a patch of eosinophilic hyaline swelling of collagen surrounded by a granulomatous reaction in which the cells are mainly lymphocytes, plasma cells, macrophages and fibroblasts. In both the subcutaneous nodule and Aschoff body, the hyaline change in collagen appears to be due to permeation by plasma proteins exuded from the small vessels. Various erythematous skin rashes may occur, the commonest being *erythema marginatum*, and in some cases there may be a mild *encephalitis*.

Clinical features

Rheumatic fever develops usually 2 to 4 weeks after a streptococcal sore throat. Symptoms may be mild, but there is usually fever, tachycardia, malaise and arthralgia flitting from joint to joint; the affected joints are sometimes swollen. The subsequent course depends on the degree of cardiac involvement: the most serious effect at this stage is on the myocardium, and various degrees of acute heart failure are observed. Signs of acute pericarditis usually appear later in the acute illness, and the valvular lesions are undetectable at this stage, although there may be evidence of secondary mitral incompetence due to dilatation of the left ventricle, or valvular abnormalities resulting from previous attacks. Involuntary movements (*chorea*) attributable to involvement of the brain may occur during or apart from the acute illness, as may the skin rashes and subcutaneous nodules.

There is no specific test for rheumatic fever. A raised erythrocyte sedimentation rate, anaemia, slight leukocytosis and high titres of streptococcal antibodies are usually present. C-

reactive protein, one of the so-called 'acute-phase reactants', appears in the serum at an early stage but is found in many other acute illnesses.

Subclinical rheumatic fever

In some patients with a clinical history of rheumatic fever, valve lesions develop after many subsequent years of apparent good health. Examination of atrial material obtained at operation for chronic valvular disease reveals florid sub-endocardial Aschoff bodies in about 50 per cent of such cases under 50 years old, and since the Aschoff body is regarded as an indication of active rheumatism, it is clear that the disease can smoulder on subclinically for many years. In some cases, there is no history of an illness suggestive of acute rheumatic fever.

Aetiology. Most attacks of rheumatic fever occur 2 to 4 weeks after infection of the throat by β -haemolytic streptococci of Lancefield group A. There is no doubt that such throat infection is a major aetiological factor since prolonged administration of penicillin to patients who have had an attack of rheumatic fever greatly reduces the incidence of further attacks of this notably recurrent disease. Presumably the high incidence of rheumatic fever in the economically poor is due to increased likelihood of streptococcal infection.

It is far from clear how preceding infection of the throat leads to the cardiac and other lesions of rheumatic fever. The 2 to 4 weeks' delay in onset speaks against a direct infection of the heart, etc., or a direct effect of exotoxin. Streptococci have not been demonstrated convincingly in the cardiac lesions, and the failure of antibiotics to prevent the disease when first administered during the interval between the sore throat and the onset of rheumatic fever is not in keeping with simple spread of infection.

There is strong evidence that the disease has an immunological basis. Compared with those who make an uncomplicated recovery from streptococcal sore throat, patients who develop rheumatic fever have unusually high antibody levels against various streptococcal antigens, including antistreptolysin O (ASO) titre, which is widely used as a diagnostic aid. It is of particular interest that Kaplan (1961) has shown that certain antigens are shared by some types of group A streptococci and by myocardial

fibres. The serum of many patients contains auto-antibodies which react *in vitro* with myocardial fibres and in fatal cases immunoglobulin and reaction-products of complement have been shown to be bound to the surface of the myocardial cells. It thus appears that throat infection by streptococci leads, in certain susceptible individuals, to strong immunity against various streptococcal antigens; the response includes the production of antibodies which cross-react with myocardial cells, resulting in activation of complement and consequently a cytotoxic effect on the myocardium. The full explanation is, however, likely to be more complex, for rheumatic fever may follow infection by streptococci which do not appear to contain the shared antigens; nor does auto-antibody to myocardium explain the Aschoff body, which does not contain fixed immunoglobulin and is widely regarded as a lesion of connective tissue rather than of myocardial fibres. The pericarditis and endocarditis are also unexplained and it is also not known why streptococcal infections of the skin and elsewhere are not followed by rheumatic fever. Furthermore some unexplained factor must render the cardiac tissues abnormally permeable to permit the union of antibody and complement with cardiac muscle sarcoplasm, for antibody to myocardium induced experimentally in animals does not become bound to the myocardium *in vivo*.

Chronic rheumatic heart disease

Chronic rheumatic heart disease is a common sequel to acute rheumatic fever and is characterised mainly by chronic endocarditis in which overgrowth of connective tissue and irregular shrinkage leads to valvular deformities. Compensatory dilatation and hypertrophy of the chambers are found, depending on the nature of the valvular lesions. In many cases Aschoff bodies are present together with minute foci of myocardial fibrosis representing healed Aschoff bodies. Pericardial adhesions are common but

these impede cardiac action only when the fibrous thickening is unusually severe and especially when the pericardium is adherent to other mediastinal structures.

Pathogenesis of lesions. The lesions result from the healing of acute rheumatic endocarditis and fibrosis may occur in all the sites previously inflamed in the acute stage. Thus the valves affected are the mitral, aortic and tricuspid in that order of frequency, and the chordae tendineae are much thickened and shortened. Fusion of the cusps along their contiguous edges by organisation of vegetations and fibrosis leads to permanent *stenosis* of the valves. The chronic inflammatory process leads to thickening and retraction of the valve cusps which prevents efficient closure and valvular *incompetence* thus results. Both effects are frequently present.

Effects

(1) **Cardiac failure** is common, mainly as a result of the mechanical problems created by stenosed and incompetent valves. These are considered in detail below.

(2) **Thrombi** frequently form in the atrial appendages, especially in patients with mitral or tricuspid stenosis and atrial fibrillation, and may cause emboli in the lungs and systemic arteries, e.g. in the brain, with serious consequences. A rare occurrence is the formation in the left atrium of a so-called ball thrombus which lies free in the cavity and which may reach several centimetres in diameter.

(3) **Angina pectoris** occurs when the cardiac output is so reduced that coronary blood flow to the hypertrophied myocardium is inadequate.

(4) **Cardiac arrhythmias**, especially atrial fibrillation, are a common consequence and may be due to myocardial and subendocardial scar tissue.

(5) **Infective endocarditis** is particularly prone to develop on valves even slightly damaged by rheumatic fever.

capillary pressure to levels which produce pulmonary oedema.

In patients with mitral stenosis and tricuspid incompetence, any severe rise of pulmonary arterial pressure is prevented by regurgitation into the easily distensible systemic venous system via the right side of the heart; this has the effect of decompressing the pulmonary circuit and bringing on right heart failure (p. 398), which is the common mode of death in mitral stenosis.

The venous congestion of the lungs in mitral stenosis cannot be alleviated by any compensatory process. The changes (p. 230) often cause slight haemoptysis, to be distinguished from the more severe haemoptysis which results from pulmonary infarction when the heart is failing. The changes in the pulmonary vessels due to pulmonary hypertension, and the factors involved in pulmonary oedema, are dealt with on pp. 451–7.

In mitral stenosis the heart comes to have a quadrangular form, but its weight is not greatly increased. There is usually a distinct presystolic murmur over the precordium and sometimes a palpable thrill during atrial systole.

Mitral incompetence

Competence of the mitral valve depends on a complexity of factors, including the size and flexibility of the annulus and cusps, and the state of the chordae tendineae, papillary muscles and left ventricle. In consequence, the diagnosis of incompetence at necropsy is far more difficult than that of mitral stenosis.

Causes. Mitral incompetence results from the following.

1. *Rheumatic endocarditis*, where it is due to retraction of the cusps and shortening of the chordae tendineae. Pure incompetence from this cause is now quite rare; combined stenosis and incompetence is more common, and is sometimes a result of the surgical treatment of mitral stenosis.

2. *Myocardial ischaemia*. Mitral incompetence from this cause is now relatively common. It may develop suddenly and in severe form from rupture of papillary muscles involved in myocardial infarction, or insidiously from ischaemic fibrosis of the papillary muscles.

3. *Myxoid degeneration of the cusps*, causing

a 'floppy' mitral valve, occurs usually in old people. Disintegration of the collagen of the valve cusps, accompanied by increase in basophilic ground substance (p. 275), results in weakness and stretching of the cusps. Minor degrees are common but without effect: if the changes are severe, the cusps are liable to prolapse during ventricular systole with consequent regurgitation of blood.

Similar changes occur rarely in young people, sometimes as a complication of Marfan's syndrome.

4. *Changes in the annulus*. Closure of the mitral valve is dependent on the size and shape of the mitral ring during ventricular systole. Regurgitation due to stretching of the ring is a regular feature of the dilatation of the failing left ventricle, and occurs also in some cases of Marfan's syndrome. Senile calcification of the ring can also cause incompetence.

Effects. Mitral incompetence results in compensatory dilatation and hypertrophy of the left ventricle. From the outset, the left atrium is dilated and the pulmonary circulation congested. Increased work is thus thrown on the right ventricle, which also becomes hypertrophied. The subsequent effects are similar to those occurring in mitral stenosis. Mitral incompetence is associated with an apical murmur throughout, or late in, ventricular systole.

Aortic stenosis

Except in countries where rheumatic heart disease is still prevalent, aortic stenosis alone is probably as common as isolated mitral stenosis, and most cases are probably non-rheumatic. The commonest form is **calcific aortic stenosis** without evidence of rheumatism, and many cases appear to result from the chronic fibrous thickening and calcification of a congenitally bicuspid valve (Fig. 15.27), so that aortic stenosis develops in middle age. The condition occurs also in old age, but usually without the underlying bicuspid abnormality (Fig. 15.28).

Fibrosis and calcification occur in the connective tissue of the cusps, which become hard and rigid. Irregular calcified nodules commonly project from the upper surface of the thickened cusps. Stenosis is often very severe and, as with mitral stenosis, it is surprising how small an orifice is compatible with life.



Fig. 15.27 Calcific aortic stenosis arising in a congenitally bicuspid valve. $\times 1.8$.



Fig. 15.28 'Senile' calcific stenosis in an aortic valve without obvious predisposing abnormality. $\times 2$.

Turbulence of blood flow may account for those cases with a bicuspid aortic valve, but the aetiology in other cases is unknown.

Congenital aortic stenosis usually becomes apparent in infancy. The valve cusps are replaced by a single diaphragm-like membrane with a central or eccentric hole. Obstruction of the left ventricular outflow by a sub-valvular membrane is another rare abnormality.

Effects. Pure aortic stenosis produces hypertrophy of the left ventricle. When the valve is both stenotic and incompetent, the ventricle

dilates and hypertrophies; the cavity becomes lengthened and more pointed. Later, the hypertrophied muscle may fail and then dilatation is the prominent feature. The passage of the blood during systole through the narrow aortic orifice produces a loud ventricular systolic murmur audible over the base of the heart, and a systolic thrill. The pulse pressure is characteristically low, and some patients die suddenly.

Aortic incompetence

This results from dilatation or distortion of the root of the aorta, from contraction or stretching of the cusps, or from a combination of both types of change. In rheumatic aortic incompetence, the cusps are thickened and contracted. In syphilitic aortitis, dilatation of the root of the aorta prevents complete closure of the valve cusps, and the resulting haemodynamic disturbance probably explains the stretching, thickening and distortion of the cusps (Fig. 15.18). Aortic incompetence occurs in some cases of Marfan's syndrome, due to weakness and stretching of the root of the aorta, and rarely to myxoid degeneration of the cusps. Another rare cause is the aortitis of ankylosing spondylitis: the aortic root is distorted and there is also damage to the cusps. Congenitally bicuspid aortic valves may also be incompetent.

Effects. Aortic incompetence is associated, from the outset, with compensatory dilatation of the cavity of the left ventricle, accompanied by hypertrophy of the wall. The internal length of the cavity may reach 12 cm or more (normal 8–9 cm), resulting in twice the normal capacity. This indicates a very gross regurgitation. Thick white collagenous patches often develop on the mural endocardium beneath the incompetent valve and are known as 'jet lesions', being attributed to forceful reflux of blood during diastole. So long as the mitral valve is competent the effects of the aortic lesion may not extend backwards to the lungs and venous system. But as the enlargement of the left ventricle progresses, and especially as the muscle fails, the muscular ring round the mitral orifice becomes stretched and secondary mitral incompetence results; the effects of the latter lesion then become superadded. Incompetence of the mitral valve from rheumatic endocarditis may be present as a concomitant lesion, and such a combination causes most striking enlargement of

the heart—the so-called *cor bovinum*, which may weigh more than a kilogram. The pulse in aortic incompetence is characteristic: there is a large systolic wave owing to the increased output by the left ventricle, and this is followed by a rapid fall. The pulse has a bounding and collapsing character—the so-called ‘water-hammer pulse’. A ventricular diastolic murmur, corresponding with the regurgitation, is usually audible over the base of the heart. There is a risk of sudden death in cases of aortic incompetence.

Tricuspid valve

Tricuspid stenosis is usually due to rheumatic endocarditis, and nearly always accompanied

by mitral and aortic valve lesions. It is usually less severe than mitral stenosis (Fig. 15.25).

Tricuspid incompetence, most often secondary to cardiac failure, produces enlargement of both right atrium and right ventricle, while tricuspid stenosis affects mainly the atrium.

Effects of valvular lesions on right side of heart. Tricuspid stenosis restricts right ventricular filling and this diminishes the likelihood of developing pulmonary oedema from the mitral stenosis, which is almost invariably present. **Pulmonary stenosis** causes hypertrophy of the right ventricle. It is very rare, but occurs as a congenital abnormality, or in the carcinoid syndrome (Fig. 15.33, also p. 649).

Infective endocarditis

In this condition, thrombi containing micro-organisms form on the endocardium, usually on the heart valves, and cause damage to the valve cusps, embolic phenomena, toxæmia and sometimes glomerulonephritis. If untreated, it is virtually always fatal, but antibiotic therapy has greatly reduced the mortality.

Traditionally, infective endocarditis is classified into *acute* and *subacute* types, the latter being a prolonged illness due to organisms of relatively low virulence and the former acute and due to virulent pyogenic bacteria. Because of the differences in the clinical pictures, the distinction between the two conditions is still valid and important, but the pattern has been somewhat altered by the widespread use of antibiotics and of drugs which depress the resistance to infection. Cases now occur which are intermediate between the two conditions, and moreover the use of the latter drugs, the development of cardiac surgery and self-administration of drugs intravenously by addicts have resulted in cases of endocarditis due to a wider range of ‘opportunistic’ micro-organisms, including fungi.

Aetiology

The factors which predispose to infective endocarditis are the occurrence of bacteraemia, lesions of the heart which favour the settling

and survival of bacteria, and depression of the general defences against infection.

Bacteraemia commonly occurs transiently and silently following tooth extractions, tonsillectomy or even vigorous chewing in the presence of periodontal infection. In bacterial endocarditis arising from such causes the bacteria responsible are usually various non-haemolytic streptococci, normally of low virulence, which colonise the mouth and were formerly classified as *Streptococcus viridans*. Bacteraemia results also from operations on the gastro-intestinal or infected urinary tracts, or even from such minor urological procedures as catheterisation and cystoscopy, and endocarditis due to *Esch. coli*, enterococci, bacteroides, anaerobic streptococci, pseudomonas, gonococcus, etc., can occasionally result.

Obvious septic lesions, such as boils, carbuncles, bacterial pneumonias and infections of the urinary, gastro-intestinal or biliary tract, are associated with occult bacteraemia, septicaemia and pyaemia. Drug addicts who use intravenous injections are also prone to infective endocarditis due to various organisms, including particularly *Staphylococcus aureus*, *Staph. epidermidis* and fungi (usually candida or aspergillus). There is also a risk of bacteraemia from indwelling venous catheters and cardiovascular surgery.

Predisposing cardiac lesions. Permanent

structural abnormalities of the valves of the heart predispose to infective endocarditis. Rheumatic and syphilitic valvular disease were formerly common precursors, but their incidence has declined in many countries, and congenital defects, such as bicuspid aortic valve have become relatively important, as also have calcific aortic stenosis and minor degrees of fibrous thickening and distortion of the valve cusps of uncertain aetiology.

It is likely that distorted valves are subject to recurrent minor trauma with loss of endothelium and that circulating bacteria stick to platelets and fibrin deposited on such lesions and become incorporated into small mural thrombi. It had been shown experimentally that when minor vascular lesions are caused by abrasion of the endothelium, intravenously-injected bacteria settle and persist in the small thrombi forming on such lesions. It appears that thrombi protect bacteria from the host's defence mechanisms, for in subacute infective endocarditis the micro-organisms survive and multiply in spite of the presence of antibodies capable of destroying any which escape into the bloodstream.

Infective endocarditis tends to develop in endocardial abnormalities at sites where the flow of blood is rapid: in addition to the heart valves, it develops also on the walls of inter-ventricular septal defects, and the only common site in the atria is on the posterior wall of the left atrium in cases of mitral incompetence, where a jet of blood from the leaking valve impinges during ventricular systole. Similarly, it may develop on the chordae and papillary muscles of a stenotic mitral valve and on the interventricular septum below an incompetent aortic valve. In these instances, it is likely that continuous minor physical trauma of the endothelium and consequent platelet deposition is caused where a rapid jet of blood strikes the endocardial surface. Infected thrombi similar to those of bacterial endocarditis occur also in the aorta above a stenotic aortic valve, in patent ductus arteriosus and in coarctation of the aorta, and it may be that localisation of bacteria depends on the recurrent deposition of platelets over a long period, sooner or later providing opportunity for colonisation by bacteria which have gained entrance to the bloodstream.

It should be added that in approximately 50

per cent of cases of infective endocarditis of the heart valves no predisposing local lesion can be detected, and this is particularly common in cases caused by *Staph. aureus* and other highly pathogenic bacteria.

Impaired defence mechanisms. Cytotoxic drugs used in the treatment of patients with cancer, renal transplant recipients, etc., result in depression of immune responsiveness and of production of polymorphs and monocytes, and thus predispose to infections of various sorts. Such therapy increases the risk of infective endocarditis, and the range of micro-organisms responsible includes opportunistic pathogens which are normally of low virulence, as well as virulent bacteria. Such therapy also alters the course of infective endocarditis, which tends to progress rapidly, even when organisms of low virulence are responsible.

Subacute infective endocarditis

This is usually caused by organisms of relatively low virulence, the most common being non-haemolytic streptococci from the mouth (see above), *Staph. epidermidis*, enterococci, *Haemophilus influenzae*, *Esch. coli*, fungi and *Coxiella burnetii*. Periodontal infection is the commonest source of the bacteraemia, but in many cases the portal of entry is not apparent. In well over half the cases there is a preceding valvular abnormality. The condition was formerly seen most commonly in children and young adults with chronic rheumatic valvular disease, but the age incidence in developed countries has risen and the mean age is now approximately 50 years.

Naked-eye appearances. *The affected valve(s)* may show evidence of chronic rheumatic endocarditis or other acquired or congenital abnormality. The vegetations are much larger, softer and more crumbling than those in rheumatic fever (Fig. 15.29). In post-rheumatic cases the mitral valve alone, or both the mitral and aortic valves, are usually involved: the aortic valve alone is involved in about 10 per cent of cases. The vegetations tend to spread on the endocardial surface and very frequently develop on McCallum's area in the left atrium, which frequently is the site of previous rheumatic lesions (Fig. 15.29). The lesions are not so rapidly destructive as in acute infective endocarditis (see below), but nevertheless cause



Fig. 15.29 Infective endocarditis of the mitral valve. The vegetations on the valve cusps are much larger than those in rheumatic endocarditis (cf. Fig. 15.22). In this instance, vegetations have formed also on the posterior wall of the left atrium. $\times 0.8$.

gross valvular injury leading eventually to heart failure. Fragments of the vegetations also break away and produce embolic lesions.

Other organs. The skin often has a brownish colour and clubbing of the fingers is usual. Large emboli cause infarction in various organs, including the brain, kidneys, spleen and intestine. The emboli seldom cause suppurating infarcts, probably owing to the high titres of antibodies which develop in the blood in the course of the prolonged infection. Petechiae found in the skin, beneath the nails and in the conjunctiva and retina, may be due to embolism (Fig. 15.30). The renal complications are described on p. 831. They include focal glomerulonephritis, which is common in cases of some months duration and is responsible for haematuria, and also diffuse glomerulonephritis which may bring about renal failure. As a rule the spleen is enlarged, sometimes markedly so.

Microscopic appearances. The vegetations consist mostly of fibrin and platelets and contain colonies of bacteria (Fig. 15.31). Polymorphs are usually scanty. The underlying cusp may be vascularised: it is oedematous, often infiltrated with polymorphs and macrophages, and there are usually foci of necrosis.

The myocardium, though grossly normal, often shows microscopic areas of infarction and inflammation, due to small coronary emboli.

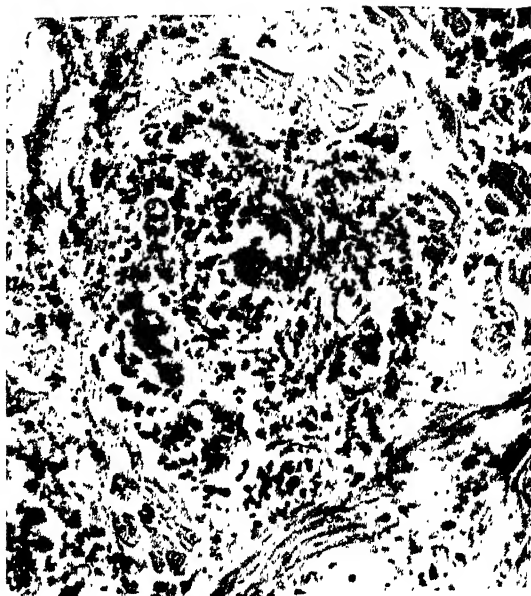


Fig. 15.30 Section through haemorrhagic spot in dermis in subacute infective endocarditis. Note small thrombosed and degenerate arteriole surrounded by leukocytes—the result of infective embolism. $\times 160$.

Clinical features. Early diagnosis of subacute infective endocarditis is very important because antibiotic therapy has reduced the mortality from virtually 100 per cent to about 30 per cent. The longer it continues, the greater the



Fig. 15.31 Colonies of *Streptococcus viridans* in a section of a vegetation in subacute infective endocarditis. (Eosin, methylene blue.) $\times 320$.

valve injury and the danger of embolic effects, etc. An irregular fever, with splenomegaly and haematuria (often only microscopic) are usually present. There may be evidence of previous valvular heart disease, and cardiac murmurs which, as the vegetations grow, alter from week to week. Various clinical effects also arise from multiple emboli, including petechial haemorrhages in the skin.

Diagnosis depends on the isolation of the causal organism by blood culture. This is often difficult, perhaps because of a high level of serum antibody, and repeated cultures may be necessary. In the absence of antibiotic therapy, persistently negative blood cultures should suggest *Coxiella burnetii* endocarditis (see below). The decision on when to stop searching for bacteria and start therapy on suspicion is often a difficult one.

Healing of the valvular lesions occurs when the organisms are eliminated by prolonged antibiotic therapy. The vegetations are removed by organisation and, depending on their size and distribution, and on the extent of valve injury, various degrees of scarring and distortion of the cusps result, with subsequent danger of heart failure.

***Coxiella burnetii* endocarditis** complicating Q fever or subclinical *Coxiella* infection is clinically and pathologically closely similar to sub-acute bacterial endocarditis. It should be suspected when blood cultures are persistently negative and the leucocyte count is normal.

Acute infective endocarditis

This progresses much more rapidly than sub-acute infective endocarditis. In untreated cases, death may occur within days from overwhelming infection or the valve lesions may cause fatal heart failure within a few weeks. In classical form, it is now caused most commonly by *Staphylococcus aureus*, and often arises as a complication of an obvious septic lesion of the skin or elsewhere. However, as explained above, it occurs also in drug addicts and various factors have resulted in conditions intermediate between subacute and acute infective endocarditis: in immunodepressed patients organisms of relatively low virulence, including fungi, can cause endocarditis of either type or of intermediate severity. The condition may supervene in patients already seriously ill with

septicaemia, and in well over half the fatal cases predisposing valvular disease cannot be detected at necropsy.

Pathological appearances. The vegetations are large and tend to break down, and the valve cusps may be largely covered by crumbling masses, which consist of layers of fibrin containing clumps of bacteria, enclosed by a zone of leukocytes, macrophages and granulation tissue. The substance of the cusps may be extensively destroyed by suppuration (Fig. 15.32): rupture, especially of an aortic cusp, may occur, leading to severe incompetence. Aneurysm of a cusp is common, the organisms causing suppuration of one side of the curtain, so that the thin tissue is stretched by the blood pressure and forms an aneurysmal bulging. The vegetations may spread also to the chordae tendineae, which may rupture. Infection may extend to the intima at the commencement of the aorta and an acute mycotic aneurysm may be formed (Fig. 15.32). The organisms may also pass directly into the adjacent heart wall, and lead to ulceration or to abscess formation. The valves most often affected are those on the left side of the heart. Involvement of the tricuspid is not uncommon, but vegetations on the pulmonary valve are rare.

The lesions are more often localised to one part of a valve or are more irregularly disposed



Fig. 15.32 Chronic rheumatic and superimposed acute bacterial endocarditis of aortic valve. The vegetations are large and irregular and there is severe damage to one of the cusps. Involvement of the root of the aorta has resulted in an acute aneurysm. $\times 0.5$.

than in the other forms of endocarditis; for example, there may be massive vegetations at the junction of two aortic cusps, the rest of the valve being free. In spite of these features, in many cases it is not possible to distinguish between acute and subacute infective endocarditis from the lesions in the heart alone.

The clinical features are those of a severe acute bacterial infection often with evidence of pyaemia or embolic phenomena, together with rapidly changing heart murmurs and the development of heart failure. In the absence of antibiotics, blood cultures are usually positive.

Other valvular lesions

Non-infective thrombotic endocarditis consists of the formation of sterile thrombotic vegetations on the heart valves, usually in a patchy fashion along the lines of closure of the cusps of the mitral and aortic valves. The vegetations are usually smaller than those in infective endocarditis and softer, more friable and less regular than those in rheumatic endocarditis. They are composed of mixed thrombus, fragments of which may break off and cause systemic embolism and infarction. In most instances, this condition is discovered at necropsy on patients who have died of cancer or other wasting diseases, and it is sometimes called *marantic* or *terminal endocarditis*. It may be associated with the venous thrombosis which occurs in some patients with carcinoma of the pancreas or of other internal organs (p. 241).

Libman-Sacks endocarditis occurs in many patients with systemic lupus erythematosus. The vegetations are sterile and are softer and more friable and usually larger than those of rheumatic endocarditis. They are also more widely dispersed on the cusps, usually affect the mitral and tricuspid valves, and may extend onto the adjacent mural endocardium or over the ventricular surface of the valve cusps (i.e. the surface away from the bloodstream).

Atheroma-like degeneration of mitral valve. Yellowish patches of thickening and degeneration similar in structure to atheroma are fairly common in the mitral valve and may be attended by fibrosis and calcification, especially at the base of the cusps. The chordae tendineae are not affected and there is seldom any effect on cardiac function.

Valvular lesions in the carcinoid syndrome. Patients with secondary carcinoid tumours in the liver may develop stenosis of the pulmonary

and, less often, the tricuspid valve. The cusps show marked fibrous thickening, have a rolled edge and are adherent along the lines of the commissures. Fibrosis may extend over the adjacent endocardium both in the right ventricle and in the wall of the right atrium (Fig. 15.33). Only trivial lesions are found in the left side of the heart, probably because of destruction of 5-hydroxytryptamine, etc., by amine oxidases in the lungs (pp. 446–7).



Fig. 15.33 Pulmonary valve in carcinoid syndrome showing great fibrous thickening of the cusp and of the subendocardial connective tissue. $\times 12$.

Disorders of the conducting system

As indicated in preceding sections of this chapter, disturbances of cardiac rhythm commonly complicate various types of heart disease. Many of them, e.g. extrasystoles, paroxysmal tachycardia and atrial fibrillation, are not usually attributable to changes in the conducting system.

The most vulnerable part of the system is the A-V bundle and its right and left branches: injury may result from the various types of myocarditis (see Fig. 15.14, p. 410), chronic myocardial ischaemia or myocardial infarction, trauma during cardiac surgery, and invasion by metastatic tumours. Bundle branch fibrosis may also arise from unknown cause. The various grades of heart block can often be ex-



Fig. 15.35 'Idiopathic' bundle branch fibrosis. *Above*, normal left bundle branch from the heart of a young man, consisting of groups of fibres lying between the ventricular myocardium and endocardial surface. *Below*, almost complete loss of left bundle branch in a case of heart block. (Some loss of fibres occurs as a normal feature of ageing.) (Professor M. J. Davies.)

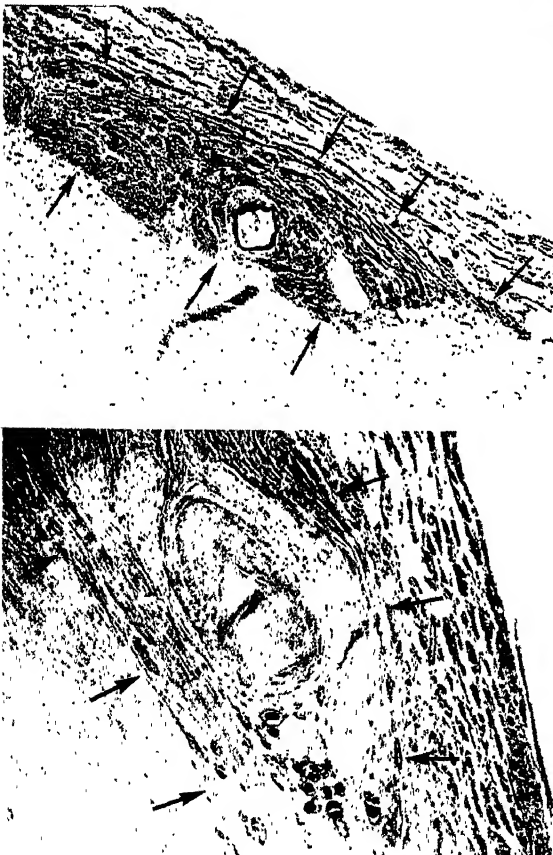


Fig. 15.34 Ischaemic fibrosis of A-V node. *Above*, normal A-V node (outlined by arrows), lying between the endocardium and central fibrous body. *Below*, ischaemic fibrosis of A-V node from a patient with heart block. (Professor M. J. Davies.)

plained by such injuries (Figs. 15.34, 15.35), but in some instances lesions have been sought in vain, while in others the finding of lesions has been associated with normal ECG patterns during life. This lack of close correlation probably reflects the large amount of work involved in thorough histological examination of the conducting system and the difficulties which arise from various artefactual changes.

Congenital Abnormalities

Little is known of the causation of congenital abnormalities, but the part played by rubella infection of the mother in the first three months of pregnancy is well established. About 10–20 per cent of the infants show serious abnormalities, of which heart disease constitutes about 50 per cent.

Normal development

A high proportion of congenital abnormalities of the heart result from defects or variations in the formation of the septa in the primitive heart. For details the student must consult a work on embryology and the classical studies of Maude Abbott and Helen Taussig, but it may be recalled that the heart at an early stage of development consists essentially of three chambers or parts, an atrial, a ventricular and the aortic bulb; division of each of these into two takes place separately. Of special importance in this connection is the relation of the ventricular septum to the division of the distal portion of the bulb into the beginning of the aorta and of the pulmonary artery. The ventricular septum grows upwards from the apex, with a curved margin resulting from the growing folds on the anterior and posterior walls, until ultimately there is a relatively small aperture at the base. The aortic bulb undergoes division into two nearly equal parts by the formation and fusion of two longitudinal folds in its wall, and the two vessels formed must rotate spirally in order to establish their normal continuity with the ventricles. The septum of the bulb ultimately fuses with the up-growing ventricular septum, the last portion to close being represented by the membranous part of the septum. Important abnormalities occur in connection with the growth of these two septa. It is to be borne in mind that the positions of the semilunar valves do not correspond exactly with the junction of the primitive ventricle and the aortic bulb. This is especially the case on the right side, where the lower part of the bulb becomes the upper part of the right ventricle or conus, and, as we shall see, this part is sometimes abnormally narrow.

While some of the anomalies are incompatible with extra-uterine life, in many the circulatory dynamics are such that the patients may survive birth for varying periods of time. With the diagnostic methods of cardiac catheterisation and angiocardiology, successful surgical cure or alleviation of many of the conditions can be effected. It is convenient to divide the anomalies into those which produce *cyanosis* and those which do not. The cyanosis is produced by admixture of a relatively large amount of reduced haemoglobin from the systemic

venous blood, with the oxygenated blood leaving the heart, i.e. a venous-arterial shunt exists. The resulting unsaturation of the arterial blood leaving the heart leads to *compensatory* rise in the red cell count, which makes cyanosis more prominent. Later, changes in the pulmonary vessels occur (p. 455) and the heart begins to fail. Cyanosis may then increase owing to impaired oxygenation of the blood by the lungs.

Cyanotic group

Malformations in connection with the aortic bulb—pulmonary and aortic stenosis. The commonest of these result from an unequal division of the bulb. Most frequently the septum is pushed to the right, so that the aorta is abnormally large and arises partly from left and partly from the right ventricle, there being usually also a defect in the ventricular septum. The result is pulmonary stenosis or obstruction, but the site of the narrowing varies. Sometimes the pulmonary artery is small, the division of the bulb being markedly unequal and occasionally the small pulmonary artery is completely obliterated. In other cases the narrowing is mainly at the valve, the cusps sometimes being partly fused to form a thickened diaphragm with an aperture of varying size. More rarely, there is a narrowing of the part of the right ventricle below the valve, that is, the part which is derived from the bulb. All these abnormalities interfere with the flow of blood into the pulmonary artery, and lead to a varying degree of hypertrophy of the right ventricle. Part of the blood from the right ventricle passes through the aperture in the interventricular septum and then into the aorta, and after birth the ductus arteriosus usually remains open and the lungs receive part of their blood supply through it. The foramen ovale also remains open and may be very large.

The commonest anomaly of this group and one which is amenable to surgery is the **tetrad of Fallot**. In this there is obstruction in the outflow tract of the right ventricle, usually from stenosis of the pulmonary valve, though the obstruction may be in the infundibular part of the right ventricle. This results in right ventricular hypertrophy and the pressure in this chamber is raised so that some of the reduced blood in the chamber is shunted through a high interventricular septal defect into the aorta, which, in addition to receiving the oxygenated blood from the left ventricle, partially overrides the septal defect and is thus in communication with the cavity of the right ventricle. All degrees of severity exist in the stenosis of the right ventricular outflow, the size of the septal defect and the dextraposition

of the aortic root. In extreme cases the pulmonary orifice and artery may be atretic and blood reaches the lungs from the aorta through a patent ductus arteriosus.

In about 25 per cent of cases of Fallot's tetrad, there is a right aortic arch.

Eisenmenger's complex. In this there is a strong resemblance in the gross morphology of the heart to that just described, but there is no obstruction to the outflow from the right ventricle. The pressure gradients across the high interventricular septal defect are such that little right-to-left shunting of blood, and hence little cyanosis, occurs at first. Later, with the onset of pulmonary hypertension and changes in the pulmonary vessels, overt cyanosis occurs, partly from admixture cyanosis and partly from faulty oxygenation of the blood by the lungs.

Transposition of the great vessels. A curious anomaly results from failure of the proximal aorta and pulmonary artery, formed by division of the aortic bulb, to undergo the rotation necessary for the establishment of their correct relationships with the ventricles. In consequence, the aorta arises from the right ventricle and the pulmonary artery from the left. While such a condition alone is incompatible with extra-uterine life, it may sometimes be compensated, for a time, by persistence of the ductus arteriosus, patent foramen ovale or a defect of the interatrial or interventricular septum; often these defects are present in combination. In this condition, the chief difficulty is not the volume of blood reaching the lungs but the effectiveness of the mechanism allowing oxygenated blood to reach the systemic circulation. Hence the greater the volume of the shunt, the better the admixture of arterial blood to venous blood and the less marked is the cyanosis.

Truncus arteriosus. In this the arrangement of the heart and emergent arteries resembles that in elasmobranch fishes in which the aorta and the pulmonary arteries arise from a common stem vessel. The pulmonary arteries may be replaced by enlarged bronchial arteries. The truncus arises from both ventricles, overriding a ventricular septal defect. Sometimes the septum may be missing so that a single ventricular cavity exists. Defects of the interatrial septum are also common.

Single ventricle with a rudimentary outlet chamber. In this condition, a single ventricle provides blood to both the aorta and pulmonary artery, which may arise separately or from a rudimentary outlet chamber. The interatrial septum may or may not develop normally, resulting in cor binatrium triloculare or cor biloculare respectively.

Tricuspid atresia. This is associated with defective development of the right ventricle which in extreme cases is virtually absent. Blood passes from the right to the left atrium through a defect in the interatrial

septum. The pulmonary artery is small, arising from the underdeveloped right ventricle. In some cases the vessel is atretic or occupies an abnormal position. Usually blood reaches the lungs from the aorta by a patent ductus arteriosus.

Aortic atresia. In this rare condition the aortic orifice is hypoplastic, the ascending aorta hypoplastic or atretic and the left ventricle poorly developed or absent. Circulation of blood is maintained by shunting of oxygenated blood from the left atrium into the right atrium and thence to the right ventricle and pulmonary artery. From this the aorta is filled via a patent ductus arteriosus.

Pure pulmonary stenosis. Here the course of the circulation is essentially normal but sometimes there is a patent interatrial septum. The lesion is a stenosis of either the pulmonary valve or the infundibulum of the right ventricle. The right ventricular myocardium is hypertrophied and able to force the blood to the lungs past the obstruction. If the interatrial septum is intact, cyanosis is not necessarily present; if there is interatrial communication a right-to-left shunt may be established with consequent cyanosis.

Anomalies of the venous return. These may involve the systemic or the pulmonary veins and vary greatly in detail. The superior and/or the inferior vena cava may open into the left atrium, thus shunting reduced systemic venous blood into the arterial side of the systemic circulation. In other cases, some of the pulmonary veins open into the right atrium: this results simply in an excessive amount of oxygenated blood being pumped around the pulmonary circulation and cyanosis will not occur.

Acyanotic group

Aortic valve stenosis and subaortic stenosis. Apart from these localised abnormalities, the heart is normal. Another isolated abnormality here is **bicuspid aortic valve**, which may later become the site of infective endocarditis or calcific stenosis.

Patent ductus arteriosus. While it will be appreciated from the foregoing description that this may coexist with many other anomalies, patency of the ductus may be the only abnormality present and closure by surgery restores the circulation to complete normality. Failure to close the ductus leads eventually to heart failure or the development of infective 'endocarditis' (endarteritis) at the site of the ductus. In a few cases there is associated pulmonary hypertension and in some the direction of blood flow in the ductus may be reversed so that unoxygenated blood passes from the pulmonary artery into the ductus and aorta distal to the ductus, usually immediately beyond the origin of the left subclavian artery. Such a patient may thus have a cyanotic tinge in the nailbeds of the toes but not in those of the hands.

Interatrial septal defect. This is one of the commonest congenital malformations of the heart. Even when the defect is large, it appears to have little effect on the circulation. Rarely a piece of detached thrombus, e.g. from the leg veins, passes from the right atrium through the defect to reach the left atrium and cause *crossed* or *paradoxical embolism*. While probe patency of the foramen ovale is very common in normal hearts (25 per cent approximately), the important malformations are of three main types: persistent ostium primum, ostium secundum and persistent atrio-ventricularis communis. In this last condition, there is often fusion of the tricuspid and mitral valves to form a common atrio-ventricular valve. Lutembacher's disease consists of an interatrial septal defect with mitral stenosis.

Interventricular septal defect. A high septal defect is frequently part of another congenital anomaly, e.g. tetrad of Fallot, but an isolated high interventricular septal defect is not uncommon. *Maladie de Roger* is the name sometimes applied to an isolated defect in the interventricular septum; the size and location of the aperture varies.

Anomalies of the aortic arch. As shown by Blalock, these are common in association with tetrad of Fallot (see above), but as isolated anomalies they rarely cause symptoms. When, however, a vascular ring is formed around the trachea and oesophagus by a right aortic arch and left descending aorta together with a persistent ductus arteriosus, ligamentum arteriosum or an anomalous left subclavian artery, pressure effects, mainly on the trachea, may result. A double aortic arch may give similar symptoms.

Coarctation (stenosis) of the aorta. Slight narrowing of the aorta between the left subclavian artery

and the orifice of the ductus arteriosus, i.e. in the interval where the two main streams of the fetal circulation cross, is not very uncommon. The stenosis is rarely marked, but it may be severe and all degrees of narrowing up to complete atresia of the aorta at this point have been recorded. With major narrowing, an extensive collateral system from the carotids and subclavians links the aorta above and below the narrowed segment. The pulses in the lower limbs are poor as compared with those of the upper. Hypertension develops and death is likely to ensue from cardiac failure, cerebral haemorrhage or less commonly from local complications associated with the constricted site, e.g. aneurysm or rupture of the aorta. Coarctation of the aorta may be associated with other congenital abnormalities, but frequently it is the only abnormality present and, moreover, it is one that can be cured by surgery. The condition is distinctly commoner in the male sex.

Ebstein's disease. In this condition there is downward displacement of the tricuspid valve so that the upper part of the right ventricle comes to be a functional part of the right atrium. The course of the circulation is normal.

Other abnormalities of the valves. Sometimes there is excess or deficiency in the number of the cusps of the semilunar valves; occasionally there are four cusps, usually somewhat unequal in size, but, as a rule, there is no interference with the efficiency of the valve. There may, however, be only two cusps, usually in the aortic valve. One cusp is usually larger than the other and often shows evidence of fusion of two cusps. Such bicuspid valves tend to develop calcific aortic stenosis (p. 418) and also bacterial endocarditis. Very rarely cases have been recorded in which two mitral valves have been present.

Diseases of the Pericardium

Pericarditis

Inflammation of the pericardium can result from bacterial and viral infections, or as a complication of myocardial infarction, and is also a feature of acute rheumatic fever and of uraemia.

Classification. Pericarditis may be classified according to its cause and may be acute or chronic. Acute cases are usually fibrinous and are divided into those with effusion (which may be serous, haemorrhagic or purulent), and those without. Some chronic cases are classified according to their effects on cardiac function (e.g. chronic constrictive pericarditis).

Causes

Pyogenic infection. Acute pericarditis may be the result of invasion by organisms from a lesion in the vicinity, such as empyema, suppurating mediastinitis, or any ulcerating tumour, e.g. of the oesophagus. In some cases, however, infection is by the bloodstream in the course of septicaemias. Suppurative pericarditis is produced chiefly by pneumococci, streptococci and staphylococci; and infection by the last may be secondary to small abscesses in the heart wall.

Tuberculosis. The pericardium is sometimes infected by lymphatic spread from an upper

mediastinal lymph node in primary tuberculosis, or by direct spread from the pleura or lung in reinfection tuberculosis.

Coxsackie virus pericarditis, sometimes accompanied by myocarditis, occurs in outbreaks among infants and also affects adults (p. 410).

Non-infective. A sterile pericarditis commonly occurs in the later stages of an acute attack of rheumatic fever and may gravely impair the heart's action. In uraemia, pericarditis is a common late event, and appears to be due to vascular or metabolic disturbances. Acute fibrinous pericarditis usually accompanies myocardial infarction and is often more extensive than the infarct. Pericarditis is a feature of polyserositis (Pick's or Concato's disease, p. 658), in which great thickening of the subserous fibrous tissue occurs.

Naked-eye appearances

Fibrinous pericarditis is found in rheumatism, uraemia, myocardial infarcts and some infective cases. The exudate usually appears first posteriorly round the large vessels at the base of the heart as an opaque, dull and roughened layer, and when it becomes abundant it forms a rough covering to the heart with irregular projections (Fig. 3.21, p. 66), giving the so-called 'bread and butter' appearance.

Pericardial effusion, sometimes exceeding a litre, is usually accompanied by fibrinous pericarditis. The effusion may be serous in rheumatic fever or myocardial infarction, haemorrhagic in tuberculosis, uraemia, myocardial infarction and infiltration by carcinoma, and purulent following invasion by pyogenic bacteria.

The ordinary sequel to pericarditis is organisation of the deposited fibrin, and adhesions ultimately form with partial or complete obliteration of the pericardial sac. Sometimes, especially in rheumatic cases, there may be repeated attacks, and great thickening of the pericardium may result. Adherent pericardium may contribute to the development of cardiac hypertrophy. Pericardial adhesions are commonly found at necropsy and often the cause is not apparent.

Slightly thickened patches of opaque and whitish appearance in the epicardium are known as 'milk spots'. They occur especially over the anterior surface of the right ventricle and the apex of the left

ventricle, and occasionally a large area of opacity is present. They are common in hypertrophied hearts and occasionally fibrous adhesions are present over an area of thickening. Milk spots are of no clinical significance.

Tuberculous pericarditis. At an early stage the changes may resemble an ordinary fibrinous pericarditis and its real nature may be discovered only on microscopic examination. In other cases the pleura may contain caseous material and this is usually followed by much thickening of the layers of the pericardium, and sometimes by calcification. In some cases at an early stage there is an abundant exudate, both fibrinous and fluid, and it may be heavily bloodstained. Ultimately the sac may be enormously distended. Tubercle bacilli are sometimes present in very large numbers in the exudate.

Chronic constrictive pericarditis is a rare condition of dense fibrous adhesions around the heart, usually commencing in childhood with a febrile illness and pericarditis clinically resembling rheumatism. Pericardial effusion is often followed by pleural effusion and later by absorption and healing with very dense fibrous tissue and sometimes calcification. The effect is to constrict the chambers of the heart and venae cavae openings, which interferes with diastolic filling; a marked rise of venous pressure occurs and so the effects resemble those of cardiac failure. Most cases are either of tuberculous or unknown origin. Surgical resection of the visceral and parietal layers of the pericardium gives relief of symptoms and for most patients the long-term prognosis is then good.

Effects of pericarditis

Many examples of pericarditis are not recognised during life. In acute pericarditis there may be pain in the chest or neck, and pericardial friction. Signs of pericardial effusion include enlargement of the 'heart' on percussion and radiologically, with a feeble apex beat in the normal position. Chronic constrictive pericarditis may be associated with systolic retraction of the chest wall and with increased venous pressure and ascites due to interference with filling of the heart; the pulse pressure is low and decreases on inspiration (pulsus paradoxus).

Pericardial haemorrhage

Haemorrhage into the pericardial sac, giving rise to *haemopericardium*, may be due to rupture of the heart itself following infarction, to rupture of an aortic aneurysm, most often an acute dissecting aneurysm which strips open the aortic wall to the base of the heart (p. 386), or to a stab wound involving the heart or a large vessel. When the bleeding is rapid, the pressure of the blood in the pericardial sac

interferes with the diastolic filling of the chambers. The output of blood from the left ventricle is greatly diminished, the blood pressure rapidly falls and death from heart failure results—this is known as **cardiac tamponade**.

Multiple minute haemorrhages occur into the layers of the pericardium in the various purpuric conditions. They are sometimes a prominent feature also in cases of death by suffocation.

Tumours of the Heart and Pericardium

Primary tumours of the heart are rare. *Fibroma*, *myxoma*, *lipoma*, *haemangioma* and *lymphangioma* are occasionally encountered, especially in the left atrium, the commonest being a myxomatous mass of considerable size, the so-called **cardiac myxoma**, projecting into the cavity from the margin of the foramen ovale: the commonly-associated mitral valve lesions may be due to haemodynamic or traumatic effects of the tumour. Cardiac myxoma sometimes has various unexplained effects, including weight loss, anaemia, a high ESR, serum protein disturbances, Raynaud's phenomenon and arthralgia. *Rhabdomyoma* of congenital origin occurs especially in the ventricles as multiple rounded nodules of pale and somewhat translucent tissue. It consists of large branching cells in which striped myofibrils are found; the cells have a somewhat vacuolated cytoplasm and contain much glycogen. In a number of cases, the tumour has been associated with multiple discrete gliomatous growths in the cerebral hemispheres—*tuberous sclerosis* (p. 772); in some cases there have

also been malformations of the kidneys and liver, and adenoma sebaceum on the face (Bourneville's disease).

Metastatic tumours in the heart and pericardium are less uncommon than is generally realised, occurring in about 10 per cent of all fatal malignancies, secondary melanotic tumours being disproportionately numerous in relation to their total incidence. Primary carcinoma of the bronchi spreads to involve the heart more frequently than any other neoplasm (31 per cent of cases); no doubt the proximity of the primary growth is a factor in this high incidence, as direct extension readily occurs to the base of the heart and pericardium.

Neoplastic invasion of the pericardium often causes a haemorrhagic inflammatory exudate. Spread of tumour into the wall of the right atrium is liable to cause arrhythmias.

References

- Davies, M. J., Fulton, W. F. M. and Robertson, W. B. (1979). The relation of coronary thrombosis to ischaemic myocardial necrosis. *Journal of Pathology* **86**, 99–110.
- Davies, M. J., Woolf, N. and Robertson, W. B. (1976). Pathology of acute myocardial infarction with particular reference to occlusive coronary thrombi. *British Heart Journal* **38**, 658–64.
- Kaplan, M. H. (1961). Immunopathological studies in rheumatic heart disease: concept of autoantibodies to heart in rheumatic fever and post-comisurotomy syndrome. In *Inflammation and Diseases of Connective Tissue*, pp.108–18. Edited by L. C. Mills and J. H. Moyer. Saunders, Philadelphia and London.
- Lyampert, I. M. and Danilova, T. A. (1975). Immunological phenomena associated with cross-reactive antigens of micro-organisms and mammalian tissues. *Progress in Allergy* **18**, 423.

Further Reading

- Abbott, M. E. S. (1954). *Atlas of Congenital Cardiac Disease*, pp. 62. American Heart Association, New York.
- Crawford, Sir Theo. (1977). *Pathology of Ischaemic Heart Disease*, pp.170. Butterworth, London and Boston.
- Hudson, R. E. B. (1965–71). *Cardiovascular Pathology*, Vols. 1–3. Edward Arnold, London.
- Paul Wood's Diseases of the Heart and Circulation* (1968). Various Authors. 3rd end., pp. 1164. Eyre and Spottiswoode, London.
- Pomerance, Ariela and Davies, M. J. (Eds.) (1975). *The Pathology of the Heart*. Blackwell Scientific, Oxford and Melbourne.
- Taussig, H. B. (1960). *Congenital Malformations of the Heart*. 3rd edn, pp. 204 and 1049. Harvard University Press, Cambridge.

Respiratory System

Introduction

The primary function of the respiratory system—oxygenation of the blood and removal of carbon dioxide—requires that air be brought into close approximation with blood. Accordingly, **the respiratory tract is particularly exposed to infection**, both by microbes in the inspired air and by spread downwards of the bacteria which commonly colonise the nose and throat. Another important hazard is presented by **inhalation of pollutants** contributed to the air we breathe in the form of dusts, smokes and fumes, a particularly important example being cigarette smoke. These pollutants are responsible for the high incidence of chronic bronchitis and chronic lung disease and also bronchial carcinoma in many parts of the world. Thirdly, the lungs are the only organs, apart from the heart, through which all the blood passes during each circulation: accordingly, cardiovascular diseases which disturb pulmonary haemodynamics are likely to have serious secondary effects on the lungs, such as pulmonary oedema, and conversely diseases of the lungs which interfere with pulmonary blood flow have important effects on the heart and systemic circulation. In short, *normal cardiac and pulmonary function are closely interdependent*.

Apart from infections, injury due to inhaled pollutants and the effects of cardiovascular disease, the respiratory tract is remarkably trouble-free, and most of this chapter will be devoted to the effects of these three hazards.

Although the respiratory tract, like other systems, is best considered on a regional basis, the continuity of the mucous membranes from nose to alveolus, and the microbial contamination of inspired air, allow ready spread of infection, and accordingly it seems appropriate to give a brief general account of the main factors

concerned in respiratory tract infections before proceeding on a regional basis.

Respiratory infections

The defences of the respiratory tract against infection have been described in Chapter 7: they include upward flow of the surface film of mucus which coats the air passages and is impelled by ciliated epithelium; the cough reflex; the secretion of IgA antibodies; and the phagocytic activity of alveolar macrophages.

Bacterial infections of the respiratory tract may be primary (i.e. occur in healthy individuals), or secondary to a large number of conditions which depress resistance. **Primary infections** are now relatively rare in many parts of the world: they include laryngeal or nasal diphtheria, bacterial pneumonia due usually to *Strep. pneumoniae*, and pulmonary tuberculosis. Other examples include pneumonic plague and anthrax pneumonia. *Primary pneumonia due to various pyogenic bacteria is, however, relatively common in infants and old people*. **Secondary bacterial infections** occur especially when the local resistance of the respiratory mucosa is lowered by various virus infections, e.g. the common cold, influenza and measles: in these conditions, bacteria growing in the nose and throat extend downwards, usually giving a mixed infection, but in hospitals and other institutions, outbreaks of respiratory virus infections may be complicated by spread of virulent pathogenic bacteria from patient to patient. Chronic liability to bacterial infections also results from persistent abnormalities of the bronchi, especially chronic bronchitis and bronchiectasis, from various debilitating and wasting diseases, and from congenital and acquired immunodeficiencies.

Virus infections. *Most acute respiratory disease seen in general medical practice every*